Title

Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era.

Journal

Animal genetics, 50(6)

ISSN

0268-9146

Authors

Raudsepp, T
Finno, CJ
Bellone, RR
et al.

Publication Date

2019-12-01

DOI

10.1111/age.12857

Peer reviewed

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

REVIEW

doi: 10.1111/age.12857

Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era

T. Raudsepp*, C. J. Finno†, R. R. Bellone†,‡ and J. L. Petersen§

*Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Research, Texas A&M University, College Station, TX 77843, USA. †Department of Population Health and Reproduction, School of Veterinary Medicine, University of California—Davis, Davis, CA 95616, USA. ‡School of Veterinary Medicine, Veterinary Genetics Laboratory, University of California—Davis, Davis, CA 95616, USA. §Department of Animal Science, University of Nebraska, Lincoln, NE 68583-0908, USA.

Summary

The horse reference genome from the Thoroughbred mare *Twilight* has been available for a decade and, together with advances in genomics technologies, has led to unparalleled developments in equine genomics. At the core of this progress is the continuing improvement of the quality, contiguity and completeness of the reference genome, and its functional annotation. Recent achievements include the release of the next version of the reference genome (EquCab3.0) and generation of a reference sequence for the Y chromosome. Horse satellite-free centromeres provide unique models for mammalian centromere research. Despite extremely low genetic diversity of the Y chromosome, it has been possible to trace patrilines of breeds and pedigrees and show that Y variation was lost in the past approximately 2300 years owing to selective breeding. The high-quality reference genome has led to the development of three different SNP arrays and WGSs of almost 2000 modern individual horses. The collection of WGS of hundreds of ancient horses is unique and not available for any other domestic species. These tools and resources have led to global population studies dissecting the natural history of the species and genetic makeup and ancestry of modern breeds. Most importantly, the available tools and resources, together with the discovery of functional elements, are dissecting molecular causes of a growing number of Mendelian and complex traits. The improved understanding of molecular underpinnings of various traits continues to benefit the health and performance of the horse whereas also serving as a model for complex disease across species.

Keywords ancient genomes, centromeres, complex traits, domestication, Mendelian traits, modern breeds, signatures of selection, Y chromosome

Introduction

The horse (*Equus caballus*, ECA) occupies a special place amongst farm animals. Since domestication about 5500 years ago (Outram *et al.* 2009; Librado *et al.* 2016; Gaunitz *et al.* 2018), horses have served humans in agriculture, warfare and transportation, and as valued companions. In modern times, they continue to interact with humans in many different ways and are an important part of the leisure industry. Humans have selectively bred horses for performance traits (speed, endurance, strength, gait), appearance (size, color, conformation) and temperament, resulting in 400–500 different breeds (Hendricks 2007; Petersen *et al.* 2013a). The derivation of breeds from selective breeding and the inclusion of only individuals with breed-defining characteristics have resulted in genomic features that vary among populations (Petersen *et al.* 2013b). Furthermore, over 130 equine hereditary traits (e.g. muscle disorders, allergies, asthma) can serve as valuable models for the study of similar human conditions (Wade *et al.* 2009; OMIA: https://omia.org/home/). Interest in economic, biomedical, evolutionary and basic science aspects of the horse combined with intense passion for advancing knowledge on equids have promoted organized studies of the horse genome, initiated at the First

Address for correspondence

T. Raudsepp, Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843-4458, USA.
E-mail: traudsepp@cvm.tamu.edu

Accepted for publication 09 August 2019

International Equine Gene Mapping Workshop in 1995 (reviewed by Chowdhary & Bailey 2003; Finno & Bannasch 2014). Whereas multiple important achievements mark the 25 years of horse genomics (reviewed by (Chowdhary & Bailey 2003; Chowdhary & Raudsepp 2006; Chowdhary *et al*. 2008; Chowdhary & Raudsepp 2008; Finno *et al*. 2009; Brosnahan *et al*. 2010; Bailey & Brooks 2013; Chowdhary 2013; Finno & Bannasch 2014; Librado *et al*. 2016), without doubt the most important milestone was the generation of the reference genome assembly (Wade *et al*. 2009), made possible through collaborative efforts of the research community. The completion of a reference genome has shaped horse genome studies during the past 10 years and, together with the ongoing revolution in genomics technologies, particularly next-generation sequencing (reviewed by van Dijk *et al*. 2014, has taken equine genomics to a new level.

This review will focus on these technology-driven achievements in the post-genome era of horse genomics: the development of new genomics tools and resources, the improvement of the reference sequence and functional annotation of the horse genome. We will discuss how the cutting-edge genomics technologies have improved understanding of the evolutionary history of the horse, the genetic makeup of individual horse breeds, and the study of Mendelian and complex traits.

The horse genome

EquCab2.0

In 2019, we celebrate the tenth anniversary of the publication of the genome sequence of the Thoroughbred mare, *Twilight* (Wade *et al*. 2009). The sequence of *Twilight* represented the first equine and Perissodactyl genome and established a reference sequence for the domestic horse (Wade *et al*. 2009). Sequencing was performed using the Sanger method with an average of 6.8-fold genomic coverage. The contiguity of the assembly was increased by the inclusion of end sequences of approximately 315 000 BAC clones from the CHORI-241 BAC library, which represents sequences from *Twilight*'s half-brother *Bravo* (https://bacpacresources.org/; Leeb *et al*. 2006). Radiation hybrid and cytogenetic maps (Raudsepp *et al*. 2008) assisted with chromosomal anchoring and orienting the scaffolds. The result was a high-quality, 2.5 Gb draft assembly, denoted as EquCab2.0, which incorporated the 31 horse autosomes, the X chromosome and the mitochondrial genome. Sequence annotation by the ENSEMBL pipeline predicted 20 322 protein-coding genes, comparable with human, mouse and other mammals. An important part of the horse genome project was the identification of over a million SNPs, which directed the development of genomic tools for mapping in the horse. The SNPs were generated from the diploid genome of *Twilight* and by partial sequencing of seven additional horses of diverse breeds. The reference assembly and the SNP map marked a turning point in horse genomics by providing resources for driving subsequent molecular, clinical and evolutionary studies in the horse (reviewed by Wade 2013; Finno & Bannasch 2014; Ghosh *et al*. 2018).

EquCab3.0

Despite its high quality, EquCab2.0 is a draft assembly and, as a product of the available technology of the time, has limitations (Kalbfleisch *et al*. 2018). The assembly contains numerous gaps, mainly in structurally complex genomic regions, which include segmental duplications and CNV sites (Doan *et al*. 2012; Ghosh *et al*. 2014). Approximately 0.2 Gb of sequence reads in EquCab2.0, mainly highly repetitive regions, remain unassembled and unassigned to chromosomes (Wade *et al*. 2009; Wade 2013). Furthermore, recent re-sequencing of the whole genome (Rebolledo-Mendez *et al*. 2015) of selected complex regions such as the MHC (Viluma *et al*. 2017) and part of the PAR (Rafati *et al*. 2016), together with transcriptome sequencing (Coleman *et al*. 2010, 2013b) and gene annotations (Hestand *et al*. 2015; Balmer *et al*. 2017; Mansour *et al*. 2017), have revealed several inconsistencies in the assembly. Therefore, taking advantage of new genomic technologies, the genome of *Twilight* was recently re-sequenced and assembled, resulting in EquCab3.0 (Kalbfleisch *et al*. 2018). The new assembly was built upon the solid foundation of EquCab2.0 (Wade *et al*. 2009), physical maps (Raudsepp *et al*. 2008) and BAC end sequences (Leeb *et al*. 2006). These were augmented with 45-fold short-read data that improved the accuracy of unique regions of the genome. Chromosome length scaffolding was achieved by including Chicago® and Hi-C proximity ligation data and 16-fold long-read Pacific Biosciences (PacBio) data.

The new assembly improved both the contiguity and composition of the horse reference genome. For example, the number of gaps was reduced 10-fold, from 55 Mb (2.2% of the genome) in EquCab2.0 to 9 Mb (0.34% of the genome) in EquCab3.0, and the number of assembled bases in the incorporated chromosomes improved from 2.33 to 2.41 Gb (3% increase). Contiguity improved nearly 40-fold and it is noteworthy that only ECA6 comprises two scaffolds; all other chromosomes comprise a single scaffold. In addition, the use of the Chromium 10X platform allowed for true haplotype phasing, so the final assembly has the most common and likely ancestral allele at each heterozygous site. Comparison of EquCab2.0 and EquCab3.0 mapping statistics for 13 previously published ancient DNA samples (Schubert *et al*. 2014; Librado *et al*. 2015, 2017; Gaunitz *et al*. 2018) showed that significantly more reads mapped to the new assembly, demonstrating its improved utility for the mapping of highly fragmented and damaged DNA samples.

Table 1 demonstrates that the size and gene content of the reference genome have marginally but consistently increased for all chromosomes. The exception is ECA5, which is smaller and has fewer genes in EquCab3.0 than in EquCab2.0, most likely owing to inconsistencies in the previous assembly. Two chromosomes, ECA12 and ECAX, increased in length by almost 4 Mb owing to incorporating previously unplaced contigs, and the number of annotated, non-coding genes has essentially increased for all chromosomes. This is probably due to improved sequence composition and better annotation pipelines, although detailed annotation of the horse genome will be the task of the ongoing Functional Annotation of ANimal Genomes (FAANG) project (described in detail below). The necessary prerequisite for FAANG is a high-quality and contiguity reference genome (Andersson *et al.* 2015), and EquCab3.0 serves this purpose. Nevertheless, EquCab3.0 is also a tribute to EquCab2.0 and testimony that the first Sanger assembly of the *Twilight* genome 10 years ago was an outstanding achievement, providing the foundation for answering genetic questions related to the horse.

Unique features of horse centromeres

Centromeres are typically not part of reference genomes because they are composed of arrays of nearly identical tandem repeats, known as satellite DNA, and are extremely difficult to assemble even with long-read sequencing technologies (Miga *et al.* 2014; Jain *et al.* 2018). An outstanding exception is horse ECA11, which has no satellite DNA and is the first sequenced example of a natural satellite-free and evolutionary 'immature' centromere (Wade *et al.* 2009; Wade 2013). This unusual and most interesting discovery led to in-depth studies of centromeres in horses and equids. Three centromere satellite families, 37cen, 2PI and EC137, have been isolated, characterized and localized in horse and equid chromosomes, revealing that in donkeys and zebras a large number of centromeres lack satellite DNA (Piras *et al.* 2010; Nergadze *et al.* 2014). With the help of chromatin immunoprecipitation sequencing (ChIP-seq) methodology, it was shown that 37cen satellite binds to the centromeric histone H3 variant CENPA protein, is transcriptionally active and probably required for centromere function (Cerutti *et al.* 2016). However, further studies of satellite-free horse ECA11 revealed that centromere is defined epigenetically by binding of the CENPA protein and not by the presence of satellite repeats (Purgato *et al.* 2015). Interestingly, the size and exact location of the approximately 100 kb (kilo-base) CENPA binding region differ between individuals, giving rise to epialleles, a phenomenon known as centromere sliding (Purgato *et al.* 2015; Giulotto *et al.* 2017). These observations were further strengthened by a study of donkey centromeres (Nergadze *et al.* 2018). The donkey genome has 16 chromosomes with satellite-free centromeres, which is perfectly compatible with genome stability and species survival. Analysis of the transmission of CENPA-binding epialleles in mules and hinnies shows that centromeric domains are inherited as Mendelian traits, although centromere sliding can happen in one generation. Overall, the discovery of satellite-free centromeres in horses and equids has provided a unique model for the study of the evolution, dynamics and molecular regulation of mammalian centromeres.

The Y chromosome

The female-based horse reference genome is incomplete because it does not include the Y chromosome. It is therefore noteworthy that, concurrently with the release of EquCab3.0 (Kalbfleisch *et al.* 2018), the first comprehensive assembly and annotation of the male-specific region of the horse Y (MSY) was published (Janečka *et al.* 2018). The 9.5 Mb assembly represents the Y chromosome of a Thoroughbred stallion *Bravo*, a half-brother of *Twilight*, thus completing the Thoroughbred-based horse reference genome. The assembly provides information about the horse Y chromosome organization, sequence classes and gene content (52 genes and 174 transcripts). Notably, the study identified a novel testis-expressed XY ampliconic sequence class *ETSTY7*, which is shared with the parasite *Parascaris* genome, providing evidence for eukaryotic horizontal gene transfer. Alignment of MSY assembly with horse, donkey and mule testis transcriptome data suggests candidate genes for stallion fertility. The MSY assembly provides a needed reference toward improved understanding for the role of the Y chromosome in equine male development and fertility.

Another keen interest in the male-specific and non-recombining Y chromosome is its inheritance exclusively through male lineages. This makes Y chromosome sequence polymorphisms excellent markers for tracing the patriline history of ancient horses and modern breeds. However, until recently, the main problem with the horse Y chromosome was the lack of sequence polymorphism (Lindgren *et al.* 2004; Wallner *et al.* 2004; Brandariz-Fontes *et al.* 2013), suggesting a limited number of patrilines in horse domestication and omitting the use of the Y to trace those patrilines. Even though one polymorphic microsatellite, *YA16*, with two alleles was detected in a few individuals of indigenous Chinese horses (Ling *et al.* 2010), no variants were found in other modern breeds. In contrast, sequencing just a 4 kb Y chromosome fragment from eight ancient horses revealed 28 segregating sites and eight haplotypes (Lippold *et al.* 2011), demonstrating considerable diversity in the ancestral horse Y and the loss of this diversity during domestication. Nevertheless, persistent search, combined with the use of high-throughput next-generation sequencing technologies, led to step-wise discovery of a limited number of variable sites also in the modern horse Y chromosome. These included two SNPs that defined six Y haplotypes in modern breeds (Wallner *et al.* 2013), and two

Table 1 Chromosome-wise comparison of EquCab2.0 and EquCab3.0.

	Size, Mb		Protein coding genes		Non-coding genes	
Horse chromosome	EquCab2	EquCab3	EquCab2	EquCab3	EquCab2	EquCab3
ECA1	185.8	188.3	1656	1683	166	705
ECA2	120.9	121.4	1062	1077	103	450
ECA3	119.5	121.4	838	883	79	391
ECA4	108.6	109.5	750	735	90	354
ECA5	99.7	96.8	1004	999	99	322
ECA6	84.7	87.2	962	988	63	326
ECA7	98.5	100.8	1236	1367	86	386
ECA8	94.1	97.6	714	773	69	397
ECA9	83.6	85.8	451	447	39	296
ECA10	84.0	85.2	1032	1146	61	346
ECA11	61.3	61.7	1086	1141	109	282
ECA12	33.1	37.0	639	738	34	177
ECA13	42.6	43.8	657	711	44	175
ECA14	94.0	94.6	665	693	63	368
ECA15	91.6	92.9	659	664	54	376
ECA16	87.4	89.0	683	713	73	299
ECA17	80.8	80.7	352	338	44	288
ECA18	82.5	82.6	422	416	62	259
ECA19	60.0	62.7	407	418	47	208
ECA20	64.2	65.3	709	738	50	262
ECA21	57.7	59.0	376	388	37	199
ECA22	49.9	50.9	533	560	53	244
ECA23	55.7	55.6	296	294	53	251
ECA24	46.7	48.3	381	453	141	298
ECA25	39.5	40.3	523	554	42	160
ECA26	41.9	43.1	221	232	19	129
ECA27	40.0	40.3	215	232	20	110
ECA28	46.2	47.3	383	388	46	189
ECA29	33.7	34.8	181	170	25	156
ECA30	30.1	31.4	171	185	20	104
ECA31	25.0	26.0	140	154	13	86
ECAX	124.1	128.2	853	821	151	371
ECAY[1]	n/a	9.5	n/a	52		n/a
Total	2440.8	2419	15 428	21 151	2055	8964

Ensembl: http://www.ensembl.org/index.html for assembly size and annotated gene content.
[1]Y chromosome data are from Janečka *et al*. (2018).

microsatellites—*YP9* in Hucul and Mongolian horses and *YN04* in a Shetland pony (Kreutzmann *et al*. 2014). This small but significant progress led to more systematic discoveries of Y chromosome variants based on WGS and MSY assembly. Fifty-three variants (50 SNPs and three indels) and 24 haplotypes were identified from a 1.46 Mb MSY sequence of 52 males of 21 different breeds (Wallner *et al*. 2017), followed by the description of another 211 variants and 58 haplotypes by screening 5.8 Mb of MSY in 130 horses of rural breeds and nine Przewalski's horses (Felkel *et al*. 2019). This is an awaited breakthrough in horse Y chromosome research and has already launched a number of studies to determine the time and cause(s) of the loss of Y variation (Wutke *et al*. 2018), as well as enabling the tracing of patrilines of modern breeds and pedigrees (Wallner *et al*. 2017; Felkel *et al*. 2018, 2019; Kakoi *et al*. 2018; Khaudov *et al*. 2018; Han *et al*. 2019).

Recent studies of ancient samples provide clues about when and why the horse Y chromosome lost its genetic diversity. A study of four MSY polymorphic markers in 96 ancient stallions from early domestication indicates that the reduction of Y diversity over time was not due to genetic drift or founder effect, but the result of artificial selection that started during the Iron Age and continued during the Roman period (Wutke *et al*. 2018). This is supported and further elaborated by a more extensive study involving 105 ancient stallions and over 1500 polymorphic MSY sites showing that Y chromosome nucleotide diversity decreased steadily during the last approximately 2000 years but dropped to present levels only after 850–1350 AD (Fages *et al*. 2019)

Genomics tools and resources

SNP genotyping array

The most impactful tool and resource for horse genomics has certainly been the reference genome. As noted, EquCab2.0 has been the critical template for the discovery

of millions of sequence polymorphisms from diverse horse breeds, leading to the development of three generations of SNP chips which have been utilized to map traits and understand breed diversity and signatures of selection. First- and second-generation DNA genotyping arrays, containing 54 602 and 74 500 SNP markers, respectively, became available in 2011 (reviewed in Finno & Bannasch 2014). A number of phenotypic traits of interest and disease traits were identified using these arrays, including Lavender Foal Syndrome (Brooks *et al.* 2010), alternate gait (Andersson *et al.* 2012), iris color variation (Mack *et al.* 2017) and ocular squamous cell carcinoma (Bellone *et al.* 2017; Tables 2 and 3). Additionally, these resources were used to identify breed specific signatures of selection (Petersen *et al.* 2013b) that aid in our understanding of the biology behind performance and other selected traits in the horse (Andersson *et al.* 2012; Petersen *et al.* 2014b).

In 2017, a third-generation SNP array was developed containing 670 805 SNP markers (MNEc670k array, Affymetrix; Schaefer *et al.* 2017). This array was designed using WGS from 156 horses representing 24 distinct breeds. Mean inter-SNP distance was estimated at 3756 bp with SNP selection aimed at tagging approximately 2 million SNPs. To date, two published studies have successfully mapped traits with the 670K array, followed by fine-mapping with WGS. The first identified a nonsense variant in *ST14* associated with Naked Foal Syndrome in the Akhal-Teke (Bauer *et al.* 2017) and the second identified genetic variants in two genes, *KRT25* and *SP6*, responsible for a curly coat in horses (Thomer *et al.* 2018).

With the recently updated reference assembly of the equine genome (Kalbfleisch *et al.* 2018), the SNP array coordinates were remapped to EquCab3.0 (Beeson *et al.* 2019). The raw reports with EquCab3.0 SNP coordinates for the MNEc670k array are hosted at https://www.animalgenome.org/repository/pub/UMN2018.1003/. Furthermore, coordinates between the two assemblies can be easily converted now at NCBI: https://www.ncbi.nlm.nih.gov/genome/tools/remap. The high-density SNP array resource is undoubtedly aiding in the mapping of other important traits and we anticipate a continued increase in the number of discoveries made possible.

WGSs of individual horses

As next-generation sequencing continues to become more affordable, WGSs of horses are being generated worldwide. At the time of this writing, 1936 public WGSs are available for the horse through the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra). This tremendous resource provides investigators with a database of horse genomes to screen potential variants, and with the continuing addition of phenotypic metadata, this resource will facilitate future studies for many years to come.

WGSs are often used to identify putative genetic variants within regions identified by GWAS. Two recent examples include the discovery of the splice site mutation in *B4GALT7* associated with dwarfism and joint laxity in Friesian horses (Leegwater *et al.* 2016). Using the first-generation SNP array (~50K markers), a region on ECA14 was initially identified by GWAS. Four dwarf cases and three unrelated controls then underwent WGS. Pooling data from the four cases and variant calling identified the putative variant in *B4GALT7*, later confirmed via Sanger sequencing and demonstrated to affect splicing through analysis of cDNA from affected horses. Notably, this is the second gene with a role in protein glycosylation in which a pathogenic mutation has been identified in Friesian horses. A nonsense mutation in *B3GALNT2* involved in muscular dystrophy with hydrocephalus in stillborn foals was discovered previously by the same group using similar techniques (Ducro *et al.* 2015). This approach was also utilized to unravel the genetic mutation responsible for both immune-mediated myositis (Finno *et al.* 2018) and non-exertional rhabdomyolysis (Valberg *et al.* 2018) in the Quarter Horse as well as congenital hepatic fibrosis in the Swiss Franches-Montagnes (Drogemuller *et al.* 2014).

Many recently discovered, potential causal genetic variants in the horse have been identified through WGS alone. Public WGS was screened for putative deleterious variants associated with stallion infertility and further evaluated in a group of 337 fertile stallions across 19 breeds (Schrimpf *et al.* 2016). A variant in *NOTCH1* (g.37455302G>A) was identified as a significant stallion fertility locus in Hanoverian stallions. Additionally, nine candidate fertility loci with missing homozygous mutant genotypes were validated. WGS has recently identified a genetic cause of occipitoatlantoaxial malformation in an Arabian horse (Bordbari *et al.* 2017) and dwarfism in a Miniature Shetland pony (Metzger *et al.* 2017). WGS data have also been used to identify novel pigmentation variants (Henkel *et al.* 2019). Finally, WGS has been used to investigate the characteristics of highly selected breeds (Metzger *et al.* 2014).

With the continued reduction in the cost of WGS, we can expect to soon see many more publicly available genomes for the horse. The beauty of this resource is that, once available, the data can be used for diverse projects worldwide.

Other tools

In the past 10 years, several other array platforms have been generated to study the horse genome. On the basis of EquCab2.0, an exon array (Doan *et al.* 2012) and two whole genome tiling arrays (Ghosh *et al.* 2014; Wang *et al.* 2014) were constructed for the discovery of CNVs. However, in a short time, these platforms have given way to more comprehensive WGSs. Likewise, cDNA and oligonucleotide microarrays, developed for gene expression studies

Table 2 Genetic variants identified for traits influencing pigmentation.

Phenotype	Gene	*Allele*	Type of variant	Chromosome	Breed	Year published	PubMed ID
Chestnut	*MC1R*	*e*	Missense	3	Many	1996	8995760
Frame overo (Lethal White Foal Syndrome, LWFS)	***EDNRB***	***O***	**Missense**	**17**	**American Paint Horse, Miniature Horse, Pinto Horse, Quarter Horse Thoroughbred, Appaloosa**	**1998**	**9530628**
Chestnut	*MC1R*	e^a	Missense	3	Black Forest	2000	11086549
Recessive black	*ASIP*	*a*	11 bp deletion		Many	2001	11353392
Cream dilution	*SLC45A2*	C^{Cr}	Missense	21	Many	2003	12605854
Sabino 1	*KIT*	*SB1*	Splicing	3	Appaloosa, Haflinger, Lipizzan, Noriker, Quarter Horse	2005	16284805
Silver (Multiple Congenital Ocular Anomalies, MCOA)	*PMEL*	*Z*	Missense		American Miniature Horse, Icelandic Rocky Mountain, Kentucky Mountain Horse	2006	17029645
Tobiano	*KIT* (proposed)	*To*	~43 Mb inversion	3	Many	2007	18253033
Dominant white	*KIT*	*W1*	Nonsense (stop-gain)	3	Franches-Montagnes	2007	17997609
Dominant white	*KIT*	*W2*	Missense	3	Thoroughbred	2007	17997609
Dominant white	*KIT*	*W4*	Missense	3	Camarillo White Horse	2007	17997609
Dominant white	*KIT*	*W3*	Nonsense (stop-gain)	3	Arabian	2007	17997609
Grey (melanoma susceptibility)	*STX17*	*G*	4.6 kb intronic duplication	25	Many	2008	18641652
Champagne dilution	*SLC36A1*	*Ch*	Missense	14	Spanish Mustang, Tennessee Walking Horse, Quarter Horse, and pony breeds	2008	18802473
Dominant white	*KIT*	*W11*	Splicing	3	South German Draft	2009	19456317
Dominant white	*KIT*	*W8*	Splicing	3	Icelandic	2009	19456317
Dominant white	*KIT*	*W7*	Splicing	3	Thoroughbred	2009	19456317
Dominant white	*KIT*	*W5*	1 bp deletion, frameshift	3	Thoroughbred	2009	19456317
Dominant white	*KIT*	*W9*	Missense	3	Holstein	2009	19456317
Dominant white	*KIT*	*W10*	4 bp deletion, frameshift	3	Quarter Horse	2009	19456317
Dominant white	*KIT*	*W6*	Missense	3	Thoroughbred	2009	19456317
Lavender foal syndrome	***MYO5A***	***LFS***	**1 bp deletion, frameshift**		**Arabian**	**2010**	**20419149**
Dominant white	*KIT*	*W12*	5 bp deletion	3	Thoroughbred	2010	https://doi.org/10.1111/j.1365-2052.2010.02135.x
Dominant white	*KIT*	*W13*	Splicing	3	American Miniature Horse, Quarter Horse	2011	21554354
Dominant white	*KIT*	*W16*	Missense	3	Oldenburg	2011	21554354
Dominant white	*KIT*	*W14*	54 bp deletion	3	Thoroughbred	2011	21554354
Dominant white	*KIT*	*W17b*	Missense	3	Japanese Draft	2011	21554354
Dominant white	*KIT*	*w17a*	Missense	3	Japanese Draft	2011	21554354
Dominant white	*KIT*	*W15*	Missense	3	Arabian	2011	21554354
Macchiato	*MITF*	*macchiato*	Missense	16	Franches-Montagnes	2012	22511888
Splashed white	*MITF*	*SW3*	5 bp deletion, frameshift	16	Quarter Horse	2012	22511888
Splashed white	*MITF*	*SW1*	Insertion 11 bp, regulatory	16	American Miniature Horse, American Paint Horse, Appaloosa, Icelandic, Morgan, Old-Tori, Quarter Horse, Shetland Pony, Trakehner	2012	22511888

Table 2 (Continued)

Phenotype	Gene	*Allele*	Type of variant	Chromosome	Breed	Year published	PubMed ID
Splashed white	*PAX3*	*SW2*	Missense	6	Lipizzan, Noriker, Quarter Horse	2012	22511888
Dominant white	*KIT*	*W18*	Splicing	3	Swiss Warmblood	2013	23659293
Dominant white	*KIT*	*W20*	Missense	3	American Paint Horse, Appaloosa, German Riding Pony, Gipsy, Noriker, Old-Tori, Oldenburg, Quarter Horse, Thoroughbred, Warmblood, Welsh Pony	2013	23659293
Dominant white	*KIT*	*W19*	Missense	3	Arabian	2013	23659293
Splashed white	*PAX3*	*SW4*	Missense	6	Appaloosa	2013	23659293
Leopard Complex Spotting (Congenital Stationary Night Blindness, CSNB)	*TRPM1*	*LP*	Insertion 1378 bp	1	American Miniature Horse, Appaloosa, Australian Spotted Pony, British Spotted Pony, Knabstrupper, Noriker, Pony of the Americas,	2013	24167615
Brindle (*Incontinentia pigmenti*)	*IKBKG*		Nonsense (stop-gain)	X	Quarter Horse	2013	24324710
Dominant white	*KIT*	*W21*	1 bp deletion, frameshift	3	Icelandic	2015	26059442
Non-dun with primitive markings	*TBX3*	*nd1*	Regulatory	8	Many	2016	26691985
Non-dun	*TBX3*	*nd2*	1609-bp and 8-bp deletion, regulatory	8	Many	2016	26691985
LP pattern modifier	*RFWD3*	*PATN1*	Regulatory	3	American Miniature Horse, Appaloosa, Australian Spotted Pony, British Spotted Pony, Knabstrupper, Noriker, Pony of the Americas	2016	26568529
Brindle 1	*MBTPS2*	*Br1*	Splicing		Quarter Horse	2016	27449517
White	*MITF*	$MITF^{244Glu}$	Missense	16	American Standardbred	2017	27592871
White leg markings	*MITF*		Regulatory	16	Menorca Purebred	2017	28084638
Dominant white	*KIT*	*W23*	Splicing	3	Arabian	2017	28378922
Dominant white	*KIT*	*W22*	Deletion 1898 bp	3	Thoroughbred	2017	28444912
Tiger eye	*SLC24A5*	*TE2*	Deletion 626 bp	1	Paso Fino	2017	28655738
Tiger eye	*SLC24A5*	*TE1*	Missense	1	Paso Fino	2017	28655738
Dominant white	*KIT*	*W24*	Splicing	3	Italian Trotter	2017	28856698
Dominant white	*KIT*	*W27*	Missense	3	Thoroughbred	2018	29333746
Dominant white	*KIT*	*W26*	1 bp deletion, frameshift	3	Thoroughbred	2018	29333746
Dominant white	*KIT*	*W25*	Missense	3	Thoroughbred	2018	29333746
Curly coat	*KRT25*	*Crd*	Missense	11	Bashkir Curly Horse	2018	29686323 29141579
Splashed white, blue eyes and deafness	*MITF*	*SW5*	63 kb deletion	16	American Paint Horse	2019	30644113
Pearl dilution	*SLC45A2*	C^{prl}	Missense	21	American Paint Horse, Lusitano, Purebred Spanish horse, Quarter Horse	2019	31006892, 30968968
Sunshine dilution	*SLC45A2*	C^{sun}	Missense	21	Standardbred × Tennessee Walking Horse cross	2019	31006892

Variants influencing pigmentation with known pleiotropic effects are in bold; details for genomic, coding and protein coordinates are in Table S1.

(reviewed by Coleman *et al*. 2013a), have been completely replaced by RNA-seq. Nevertheless, a few genomics resources from the past, such as the genomic BAC library CHORI-241 (https://bacpacresources.org/), remain in use in the post-genome era. End sequences of CHORI-241 BAC clones (Leeb *et al*. 2006) helped to validate the EquCab3.0 assembly (Kalbfleisch *et al*. 2018); BAC tiling paths are used to re-sequence complex genomic regions such as the terminal end of the pseudoautosomal region in ECAX (Rafati *et al*. 2016) and the male-specific region of ECAY (Janečka *et al*. 2018), and cytogenetic mapping of BAC clones is still a reliable method to validate copy number changes (Ghosh *et al*. 2014; Staiger *et al*. 2016a) and other large-scale structural rearrangements.

Ancient genomics and natural history of the horse

The assembly of the horse reference genome EquCab2.0 (Wade *et al*. 2009), along with success in extracting and sequencing ancient genomic DNA (Orlando *et al*. 2011), has essentially expanded our knowledge about the natural history of equids (Orlando *et al*. 2013; Jonsson *et al*. 2014), horse domestication and the dynamics of the horse genome over time, from the pre-domestication era to the present (Schubert *et al*. 2014; Librado *et al*. 2015, 2017; Gaunitz *et al*. 2018; Janečka *et al*. 2018; Wutke *et al*. 2018; Fages *et al*. 2019). The findings of ancient DNA studies were summarized and discussed in an excellent review by Librado *et al*. (2016). Therefore, here we highlight only the most notable discoveries and discuss the studies published since that review.

It is noteworthy that, with regards to ancient DNA, horses/equids have made history three times—by being the first and the oldest, and most recently, by spanning the largest timescale of ancient genome data among non-human organisms (Fages *et al*. 2019). The first successful ancient DNA extraction and analysis ever was reported in 1984 from 150-year-old tissue from the extinct quagga (Higuchi *et al*. 1984), preceding by a year the first human ancient DNA study from a 2400-year-old Egyptian Mummy (Pääbo 1985). Likewise, to date the oldest genomic DNA sample which has been successfully sequenced is that of a 700 000-year-old Pleistocene horse (Orlando *et al*. 2011, 2013), exceeding the age of the oldest hominin genomic DNA extracted from 430 000-year-old bones from Sima de los Huesos (Meyer *et al*. 2016). The current peak of ancient genomics of non-human organisms, but perhaps not the limit, is a study tracking 5000 years of horse domestication based on genome-scale data from 278 ancient animals (Fages *et al*. 2019).

Horse ancestry and domestication

Genomics-based searches into the wild ancestry of the domestic horse and the origins of horse domestication have been ongoing for decades. Yet the answers have only recently started to emerge, largely thanks to the contribution of WGS from ancient equine samples (Schubert *et al*. 2014; Librado *et al*. 2015, 2016; Gaunitz *et al*. 2018; Fages *et al*. 2019). These studies reveal that, in addition to the two extant horses, the domestic (*Equus caballus*) and the Przewalski's horse (*E. przewalskii*) (Der Sarkissian *et al*. 2015), there existed other, now extinct, horse lineages at the time of early domestication. One lineage, that was initially identified from approximately 43 000- to 5000-year-old bones from the Holarctic region (Schubert *et al*. 2014), extended to Southern Siberia (Fages *et al*. 2019). This lineage shares morphological similarities with an extinct horse described as *Equus lenensis* (Boeskorov *et al*. 2018; Fages *et al*. 2019). Recent mtDNA and Y haplotype analyses of an approximately 24 000-year-old specimen from the Tuva Republic suggest that there may have been another genetically divergent lineage of horses in Siberia and the New Siberian Islands, although the genetic contact between *E. lenensis* and this 'ghost' lineage remains unknown (Fages *et al*. 2019). Finally, Iberian samples from the third and early second millennia BCE, cluster separately from *E. caballus*, *E. przewalskii* and *E. lenensis*, have extremely divergent Y and mtDNA, and therefore suggest that there was a different, now extinct, horse lineage in Iberia during the early phase of horse domestication (Fages *et al*. 2019). Earlier it was reported that that the Holarctic horse (*E. lenensis*) contributed 12.9% to the genetic makeup of domestic horses (Schubert *et al*. 2014). However, the most recent study of 278 ancient equids and modern horses shows that none of the above extinct horse lineages contributed significantly to modern horse diversity (Fages *et al*. 2019).

Another question of interest is the genetic relationship between the two surviving horses—the domestic horse and the Przewalski's horse. Until recently, the overall consensus, based on modern and ancient WGS, was that the two are separate species, diverged approximately 45 000 years ago (Goto *et al*. 2011; Schubert *et al*. 2014; Der Sarkissian *et al*. 2015), with extensive bi-directional gene flow (Der Sarkissian *et al*. 2015; Librado *et al*. 2016). All studies agree that the domestic horse is not a direct descendant of the Przewalski's horse (Librado *et al*. 2016). These views were recently shaken by a study of over 40 ancient horse genomes from Eurasia, providing striking evidence that the Przewalski's horse is not truly wild, but rather a feral horse descended from the horses domesticated by Botai culture some 5500 years ago (de Barros Damgaard *et al*. 2018; Gaunitz *et al*. 2018). At the same time, all studied domestic horses dated from 4000 years ago to present show only 2.7% Botai ancestry, suggesting they descended from a different lineage of wild horses that subsequently went extinct (Fages *et al*. 2019).

Ancient DNA studies have also attempted to decipher, but have not completely resolved, the timeline and geography of

horse domestication. Current archeological and DNA evidence suggests multiple sites of horse domestication of which the earliest (~5000–5500 years ago) were in the Western Eurasian steppes: Botai culture in Northern Kazakhstan and the Pontic–Caspian steppe (Outram *et al.* 2009; Warmuth *et al.* 2012; Librado *et al.* 2016; Gaunitz *et al.* 2018; Fages *et al.* 2019), followed by additional candidate sites in Iberia, Eastern Anatolia, Western Iran, Levant and Eastern Europe (Hungary; Gaunitz *et al.* 2018). It is noteworthy that native breeds from the main domestication sites such as the Pontic–Caspian steppes still represent hotspots of genetic diversity for horses (Warmuth *et al.* 2011; Librado *et al.* 2016). However, even though the main candidate sites for domestication have been identified, the geographic origins of the modern domestic horse remain unknown. Given that the ancient Botai and Iberian lineages did not contribute substantially to modern domesticates and the temporal origins of the modern horse are modeled within the third and fourth millennium BCE, future studies of this timeline in other candidate regions of early domestication are needed (Fages *et al.* 2019).

Genetic cost of domestication

Domestication reduces overall fitness, known as 'the genetic cost of domestication' (Lu *et al.* 2006; Moyers *et al.* 2018). However, because of the extinction of wild horses, it has been difficult to evaluate the extent of this cost for the horse. Once again, the major breakthrough came with ancient DNA sequencing, providing a comparison with horses predating domestication (Schubert *et al.* 2014) and from early stages of domestication (Librado *et al.* 2017; Fages *et al.* 2019). Genetic changes associated with horse domestication can be summarized as follows:

1 *Extreme mitochondrial DNA (mtDNA) diversity*. Modern horses have high mtDNA diversity and lack phylogeographic mitochondrial structure, resulting in limited correspondence between mtDNA haplotypes, breeds and geography (Vila *et al.* 2001; Achilli *et al.* 2012; Librado *et al.* 2016). Sequencing horse genomes from early domestication sites show that similar mtDNA diversity was present in Scythian horses some 2300 years ago (Librado *et al.* 2017) and perhaps even earlier. Mitochondrial Bayesian Skylines reconstructed from 211 mitochondrial genomes suggest horse demographic expansion about 4500 years ago (Gaunitz *et al.* 2018; Fages *et al.* 2019). Reasons for the lack of mitochondrial structure are thought to be many, including sex-biased restocking from the wild and human management (Librado *et al.* 2016).
2 *Extreme lack of Y chromosome diversity*. Contrasting the diversity of mtDNA, the Y chromosome of domestic horses has become almost homogeneous with just a few haplotypes segregating in modern populations (Wallner *et al.* 2017; Wutke *et al.* 2018). At the same time, Y sequences of ancient Scythian horses indicate that a large diversity of domestic male founders contributed to early domestication (Librado *et al.* 2017). Thus, the observed decline in Y chromosome variation happened only in the past few thousand years, probably because of the human-mediated reduction in the stallion population size and selective breeding (Librado *et al.* 2017; Wutke *et al.* 2018; Fages *et al.* 2019).
3 *Genetic load*. As a side-effect of human-driven selective breeding, the genome of the domestic horse has an increased level of deleterious mutations compared with pre-domestication genomes (Schubert *et al.* 2014). However, as mutation load is also lower in early domesticates from ancient Sintashta and Scythia, the excess of deleterious mutations in present day horses is probably a consequence of the past 2300 years of selective breeding (Librado *et al.* 2017), whereas the most striking changes have occurred during the past 200 years (Fages *et al.* 2019).
4 *Domestication-associated genetic changes*. One of the first comparisons between ancient pre-domestic horse genomes and those of modern breeds revealed 125 candidate genes that underwent episodes of positive selection during domestication (Schubert *et al.* 2014). These include genes involved in the cardiac and circulatory system, bone, limb and face morphogenesis, brain development and behavior, and coat color (Schubert *et al.* 2014; Librado *et al.* 2016). Genetic changes during horse domestication agree with the neural crest hypothesis and involve developmental changes affecting tissues and cell types of neural crest origin (Librado *et al.* 2017). However, the genomic cost of domestication and modern breeding is best illustrated by a striking discovery of a recent comparative study of ancient and modern horse genomes, showing that early breeders managed to maintain genetic diversity for millennia after domestication, and that the genetic diversity of the modern horse has dropped by 16% during just the past 200 years (Fages *et al.* 2019). It is not a coincidence that the last 200 years also cover the time of the development of horse breeds, the establishment of studbooks and the implementation of extensive selective breeding.

Genetic makeup of modern horse breeds

Since domestication, the genetic diversity present in ancient horse populations has been exploited for selective breeding for a wide range of phenotypes. However, the creation of the about 500 specialized horse populations or breeds by intense artificial selection and the establishment of (closed) studbooks happened only during the past 100–200 years (Hendricks 2007). Owing to the wealth of available genome-wide tools and resources (see above), these breeds

can now be studied in detail for genetic makeup, signatures of selection and relatedness to other breeds. The number of publications in the field is growing almost exponentially and these include a few seminal studies encompassing a global collection of breeds (McCue *et al*. 2012; Petersen *et al*. 2013a,b; Jagannathan *et al*. 2019), and many studies specializing in a single breed or a group of related breeds, for example, the American Quarter Horse (Petersen *et al*. 2014a; Avila *et al*. 2018; Marchiori *et al*. 2019; Pereira *et al*. 2019), the Hanoverian and other German Warmblood breeds (Nanaei *et al*. 2019; Nolte *et al*. 2019), the Arabian and related Middle Eastern breeds (Almarzook *et al*. 2017; Sadeghi *et al*. 2019), and some native/primitive breeds such as Hucul and Konik (Gurgul *et al*. 2019), Chinese native horses (Zhang *et al*. 2018), Japanese native breeds (Tozaki *et al*. 2019), the Yakutian Horse (Librado *et al*. 2015) and Korean horses (Seong *et al*. 2019). This is by no means an exhaustive list and it is even longer when including breeds that have recently been studied based on a genome-wide collection of microsatellite markers, such as Estonian horse breeds (Sild *et al*. 2019) or Konik (Szwaczkowski *et al*. 2016).

Despite the large number of studies, the core findings are rather similar. Collectively, modern horse breeds are characterized by high inter-breed and low intra-breed genetic diversity (McCue *et al*. 2012; Petersen *et al*. 2013a). Genomes of modern horses show multiple regions with signatures of selective pressures on performance traits and phenotypes. Among these, the most prominent are the *MSTN* gene in ECA18 for muscle fibers in racing breeds (Petersen *et al*. 2013b; Avila *et al*. 2018), the *DMRT3* gene in ECA23 to perform alternative gaits in many breeds (Petersen *et al*. 2013b), and a region in ECA11 for body size in draught breeds and Miniature Horses (Petersen *et al*. 2013b). Clear signatures of selection are also found at known coat color loci. For example, the recessive chestnut coat color locus at *MC1R* is defined by a conserved approximately 750 kb haplotype across all breeds studied (McCue *et al*. 2012) and robust signatures of selection at the *TBX3* locus in ECA8 are found in Konik horses, known to be selected for the dun color (Gurgul *et al*. 2019). Finally, selective breeding for the desired traits in modern breeds has unwillingly introduced accumulation of deleterious mutations (Librado *et al*. 2016, 2017; Jagannathan *et al*. 2019). For example, recent WGSs of 88 horses across 25 breeds identified heterozygotes for two potentially deleterious recessive alleles: a nonsense variant in the *PALB2* gene, which is essential for mesoderm development, and a nonsense variant in the *PLEKHM1* gene, necessary for osteoclast functions (Jagannathan *et al*. 2019).

Investigating Mendelian traits

The investigation of Mendelian traits in the horse began in the early 1900s. Because of the relative ease of tracking simply inherited phenotypes across generations, some of the first Mendelian traits to be studied involved variations in pigmentation. In fact, Alfred Sturtevant, an undergraduate student working on mapping traits in *Drosophila* under Thomas Hunt Morgan, was among the first to publish on inheritance of coat color in harness horses (Sturtevant 1920). However, it was almost 100 years later when the genetic mechanism proposed by Sturtevant for the chestnut coat color was identified at the molecular level (Table 2). This and other variants identified in the early 2000s influencing pigmentation were discovered using candidate gene approaches. To date, 58 variants affecting pigmentation have been described, including 27 in the *KIT* gene that contribute to the dominant white phenotype. The identification of the majority of these pigmentation variants was made possible by the high quality of the horse reference genome sequence and available SNP array tools (Tables 2 and S1 and described above). For example, the initial horse linkage map developed by the horse genomics community (Penedo *et al*. 2005) allowed for the mapping of several pigmentation traits with pleiotropic effects, such as leopard complex spotting (Terry *et al*. 2004) and gray depigmentation (Pielberg *et al*. 2005). However, the discovery of causal variants was possible only through the availability of a reference genome that enabled both whole genome and RNA-seq data analyses to uncover large structural variants (Rosengren Pielberg *et al*. 2008; Bellone *et al*. 2013).

The first genetic disorder to be identified at the molecular level in the horse was hyperkalemic periodic paralysis, reported in 1992. Much like the early work investigating pigmentation in the horse, this was discovered by a candidate gene approach. At the time the first draft of the horse genome sequence was complete, nearly 15 years later, only nine disease causing mutations had been discovered (Tables 3 and S2). In the last several years, there has been acceleration in the discovery of causal or highly associated variants for Mendelian traits, with 14 variants reported in the last four years. This acceleration in discovery is due to advances in genomic tools and resources. DNA testing for these Mendelian traits has enabled marker-assisted selection to be utilized by horse breeders and in some cases is helping to assist in the clinical management of disease.

Genetics of complex traits

Traits considered complex are those which are not determined by one or a few genomic variants but rather by small contributions of many and perhaps hundreds of variants across the genome. These traits are generally lowly heritable, as the phenotypic outcome is determined not only by the underlying genetic variation but also by a significant impact of the environment (e.g. nutrition, training). The role of the environment complicates both the identification and understanding of the genetic components driving these

Table 3 Genetic variants underlying disease and performance traits in the horse.

Disease	Gene	Type of variant	Mode of Inheritance	Chromosome	Breed	Year published	PubMed ID
Hyperkalemic periodic paralysis	*SCN4A*	Missense	Dominant	11	American Quarter Horse and related breeds	1992	1338908
Ovotesticular disorder of sexual development (DSD)	*SRY*	Large deletion of the DNA-binding domain of the SRY gene	Y-linked	Y	Standardbred	1995	7558880
Severe combined immunodeficiency disease (SCID)	*PRKDC*	Deletion 5 bp	Recessive	9	Arabian	1997	9103416
Junctional epidermolysis bullosa (JEB1)	*LAMC2*	Insertion 1 bp	Recessive	5	Belgian and Italian draft horse	2002	12230513
Malignant hyperthermia (MH)	*RYR1*	Missense	Dominant	10	American Quarter Horse	2004	15318347
Glycogen branching enzyme deficiency (GBED)	*GBE1*	Nonsense (stop-gain)	Recessive	26	American Quarter Horse and related breeds	2004	15366377
Thrombasthenia	*ITGA2B*	Missense	Recessive	11	American Quarter Horse & Thoroughbred	2006	16407493
Thrombasthenia	*ITGA2B*	Deletion 10 bp	Recessive	11	American Quarter Horse	2007	17338169
Hereditary equine regional dermal asthenia (HERDA)	*PPIB*	Missense	Recessive	1	American Quarter Horse	2007	17498917
Polysaccharide storage myopathy (PSSM1)	*GYS1*	Missense	Incompletely dominant	10	American Quarter Horse, American Paint Horse, Appaloosa, Draft, Pony of the America, and Warmblood	2008	18358695
Junctional epidermolysis bullosa (JEB2)	*LAMA3*	Deletion 6589 bp	Recessive	5	American Saddlebred	2009	19016681
Racing distance	*MSTN*	Insertion 227 bp, regulatory		18	Thoroughbred	2010	20098749 25160752 30379863
Cerebellar abiotrophy (CA)	*MUTYH*	Regulatory	Recessive	2	Arabian, Bashkir Curly Horse, Danish Sport Horse, Trakehner, and Welsh Pony	2011	21126570 and 29103988
Foal immunodeficiency syndrome in the Fell and Dales pony (FIS)	*SLC5A3*	Missense	Recessive	26	Dales Pony and Fell Pony	2011	21750681
Androgen insensitivity syndrome (AIS)	*AR*	Regulatory	X linked	X	American Quarter Horse	2012	22095250
Myotonia	*CLCN1*	Missense	Recessive	4	New Forest Pony	2012	22197188
Permissive to gait	*DMRT3*	Nonsense (stop-gain)	Dominant	23	Numerous breeds	2012	22932389
Warmblood fragile foal syndrome (WFFS) or Ehlers–Danlos syndrome, type VI	*PLOD1*	Missense	Recessive	2	Warmblood	2015	25637337
Hoof wall separation syndrome	*SERPINB11*	Insertion 1 bp	Recessive	8	Connemara	2015	25875171
Hydrocephalus	*B3GALNT2*	Nonsense (stop-gain)	Recessive	1	Friesian	2015	26452345
Androgen insensitivity syndrome (AIS)	*AR*	Missense	X linked	X	Thoroughbred	2016	27073903
Skeletal atavism	*SHOX*	2 over lapping deletions 160 = 180 kb and 60–80 kb	Recessive	X and Y PAR	Shetland	2016	27207956
Dwarfism, Friesian	*B4GALT7*	Missense	Recessive	14	Friesian	2016	27793082
Dwarfism, ACAN-related D3*	*ACAN*	Missense	Recessive	1	Miniature Shetland	2017	27942904
Occipitoatlantoaxial malformation (OAAM)	*HOXD3*	Deletion 2.7 kb	Recessive	18	Arabian	2017	28111759
Androgen insensitivity syndrome (AIS)	*AR*	Deletion 25 bp	X linked	X	Warmblood	2017	28192783

Table 3 (Continued)

Disease	Gene	Type of variant	Mode of Inheritance	Chromosome	Breed	Year published	PubMed ID
Naked foal syndrome	*ST14*	Nonsense (stop-gain)	Recessive	7	Akhal-Teke	2017	28235824
Ocular squamous cell carcinoma (ocular SCC)	*DDB2*	Missense	Recessive	12	Belgian, Haflinger, Percheron, Rocky Mountain Horse	2017	28425625
Immune-mediated myositis (IMM/MYH1)	*MYH1*	Missense	Recessive	11	American Quarter Horse	2018	29510741
Curly coat with hypotrichosis Crd	*KRT25*	Missense	Dominant	11	Bashkir Curly Horse	2018	29686323 29141579
Curly coat without hypotrichosis	*SP6*	Missense	Dominant	11	American Bashkir Curly Horse and Missouri Foxtrotter	2018	29686323
Dwarfism, ACAN-related D4	*ACAN*	Deletion 42 bp	Recessive	1	Miniature	2018	30058072
Dwarfism, ACAN-related D2	*ACAN*	Missense	Recessive	1	Miniature	2018	30058072
Dwarfism, ACAN-related D1	*ACAN*	Deletion 1 bp	Recessive	1	Miniature	2018	30058072

Genomic, coding and protein sequence coordinates are in Table S2.

phenotypes. That said, complex traits are of significance to the industry, both economically and for animal wellbeing and include measures of athletic performance, growth and body size, and disorders such as metabolic syndrome, laminitis, equine asthma or recurrent airway obstruction (RAO) and osteochondrosis (OC), among others. Given the availability of genome-wide genotyping technologies, researchers have been working to piece together the role of genetics in these complex traits. What is known regarding the genetic basis of several of the most highly studied complex traits in horses is outlined below.

Size

Unlike many complex traits, the size of the horse is highly heritable with estimates for wither height ranging from 0.52 to 0.78 (Molina *et al.* 1999; Zechner *et al.* 2001; Suontama *et al.* 2009; Signer-Hasler *et al.* 2012). Not surprisingly, positive genetic and phenotypic correlations are reported between wither height and other growth phenotypes such as body length, heart girth and cannon bone circumference (Molina *et al.* 1999; Sadek *et al.* 2006). Providing insight into the biology underlying size, Makvandi-Nejad *et al.* (2012) employed the 50K SNP array to identify four loci that explained a majority of variation in size across-breeds. The loci identified in this work include *LCORL/NCAPG*, *HMGA2*, *ZFAT* and *LASP1*, most of which have also been shown to play a role in size and growth phenotypes in other species (Snelling *et al.* 2010; Weikard *et al.* 2010; Lindholm-Perry *et al.* 2013; Saatchi *et al.* 2014). Additional work has supported the association of *LCORL/NCAPG* (Signer-Hasler *et al.* 2012; Tetens *et al.* 2013; Staiger *et al.* 2016a; Tozaki *et al.* 2017), *ZFAT* (Signer-Hasler *et al.* 2012; Tozaki *et al.* 2017), and *HMGA2* (Frischknecht *et al.* 2015) with height in various other populations. Whereas the means by which these loci act to alter growth traits is not fully understood, a missense mutation in *HMGA2* was reported to alter DNA binding, which was attributed to reduced size in ponies (Frischknecht *et al.* 2015) and was also recently correlated with metabolic syndrome in Welsh ponies (Norton *et al.* 2019a). Functionally, variation of *LCORL* was shown to alter its expression, which may explain its role in determining an individual's size (Metzger *et al.* 2013). Interestingly, *LCORL* has also been associated with other complex disorders, including recurrent laryngeal neuropathy (Boyko *et al.* 2014) and OC (Corbin *et al.* 2012), serving as another example of the pleiotropic effects of loci involved in complex traits. As height has been associated with both OC and recurrent laryngeal neuropathy (Brakenhoff *et al.* 2006; McGivney *et al.* 2019), these associations are biologically relevant. The power of high-density genome-wide genotyping arrays has also enabled the identification of loci with more minor effects than those found by Makvandi-Nejad *et al.* (2012) or which may be breed specific (Meira *et al.*

2014b; Frischknecht *et al*. 2016; Metzger *et al*. 2018). It is important that height is treated distinctly from dwarfism, which is simply inherited in several populations including the Miniature Horse (Metzger *et al*. 2017; Eberth *et al*. 2018) and Friesian (Leegwater *et al*. 2016; Table 3).

Athletic performance

The athletic ability of a horse, whether it be jumping, racing, cutting or pulling, is clearly complex, depending upon efficient metabolic and musculoskeletal properties as well as intricate interactions and the influence of training and husbandry. Particularly in Thoroughbreds, the genetics of racing performance has been of long-standing interest. Prior to the availability of genotyping technologies, pedigree and performance data were used to examine the genetic components of phenotypic variance for Thoroughbred, Standardbred and sport horse performance (Hintz & Vanvleck 1978; Ojala *et al*. 1987; Tavernier 1990; Árnason 2001; Ricard & Chanu 2001; Langlois & Blouin 2004, 2007). Heritability estimates for racing in the Thoroughbred range greatly from nearly zero to upwards of 0.75, depending upon the specific phenotype measured (e.g. race time, race length or race winnings) and the model assumed (O'Ferrall & Cunningham 1974; Gaffney & Cunningham 1988; Williamson & Beilharz 1996; Thiruvenkadan *et al*. 2009a). As genotyping methodologies became available, regions implicated in racing performance have been identified using candidate gene (Gu *et al*. 2010; Hill *et al*. 2010a) and genome-wide association analyses (Binns *et al*. 2010; Hill *et al*. 2010b; Tozaki *et al*. 2010; Shin *et al*. 2015) and transcriptome analyses (Park *et al*. 2014), and through the identification of selective sweeps in racing populations (Moon *et al*. 2015). Genes implicated include *COX4I2* (Gu *et al*. 2010) and *PDK4* (Hill *et al*. 2010a), both involved in cellular respiration. Although these studies have uncovered evidence of genetic factors involved in the racing performance of Thoroughbreds, a single locus, *MSTN*, has been repeatedly associated with racing performance (Binns *et al*. 2010; Tozaki *et al*. 2010). In 2010, an intronic variant of *MSTN* was noted to be predictive of a horse's best racing distance (Hill *et al*. 2010b); the use of a genetic test for this variant has been adopted as a means to tailor training programs or choose matings. Since the first publication of this variant, several lines of evidence have supported the role of a SINE insertion in the promoter of *MSTN*, in high linkage disequilibrium with the intronic SNP in the Thoroughbred breed, as the functional variant (Petersen *et al*. 2014b; Santagostino *et al*. 2015; Rooney *et al*. 2018).

Outside of the Thoroughbred, the *MSTN* variant predictive of better suitability as a sprinter is highly frequent in the Quarter Horse and associated with a higher proportion of fast-twitch muscle fibers (Petersen *et al*. 2013b, 2014b). Additional studies in the racing Quarter Horse have identified other loci associated with racing performance; as a result of these works, genes involved in muscle contractility, skeletal development and neurologic function have been suggested to be associated with sprinting (Meira *et al*. 2014a,c; Beltran *et al*. 2015). In some cases, variants for racing speed are common across breeds (Shin *et al*. 2015; Pereira *et al*. 2016), fitting a hypothesis that these horses shared selective pressures for superior metabolic and musculoskeletal traits. Efforts to understand the genetic components of variation in harness racing horses suggest that their heritability is low to moderate (reviewed in Thiruvenkadan *et al*. 2009b). In the harness racing populations, similar to the role of *MSTN* in the Thoroughbreds, a single variant was reported to impact performance to the extent that the variant is nearly fixed in trotting breeds (Promerova *et al*. 2014). The gene implicated, *DMRT3*, alters motor coordination and stride length (Andersson *et al*. 2012); horses homozygous for the variant perform at a higher level than those heterozygous or wt (Jaderkvist *et al*. 2014; Jaderkvist Fegraeus *et al*. 2015). Interestingly, this variant was identified not in a study of racing performance but in a study of another complex trait in the horse—the ability to perform an alternative gait (Andersson *et al*. 2012), again demonstrating the interplay among physiological pathways. As an aside, whereas the *DMRT3* variant is deemed necessary for 'gaitedness', the basis of variations in gait is yet to be understood (Patterson *et al*. 2015; Staiger *et al*. 2016a; Fegraeus *et al*. 2017; Fonseca *et al*. 2017).

Requiring a different type of athleticism, endurance racing studied in Arabian horses identified five QTL, including genes hypothesized to be involved in neuronal and cardiac function (Ricard *et al*. 2017). As in complex traits, how the regulation of these genes or variants within them may enhance performance is an area of study. The differentially expression of microRNAs prior to and after endurance exercise is probably one means by which gene regulation is altered to allow horses to endure and excel at this type of performance (Mach *et al*. 2016).

Sport horses require yet another type of athleticism and the heritability and genetics of show jumping and dressage have been studied in several European populations. Heritability estimates, calculated for the longevity (years) of performance, have been similar across studies ranging from 0.07 to 0.18 (Ricard & FournetHanocq 1997; Braam *et al*. 2011; Seiero *et al*. 2016; Ricard *et al*. 2017). Heritability estimates for show jumping range from 0.11 (Sole *et al*. 2017) to 0.61 (Stock & Distl 2007), and in most cases are greater than estimates for dressage in the same population (Viklund *et al*. 2010; Braam *et al*. 2011), although this is not the case in the Swedish Warmblood studied by Wallin *et al*. (2003).

Finally, the use of the Illumina Equine50 SNP array to investigate the biology underlying the success of jumpers revealed a QTL explaining 0.7% of the phenotypic variance in French Warmbloods in a region including the candidate gene *RYR2* (Brard & Ricard 2015). In the Hanoverian, a

GWAS using the same genotyping platform identified six QTL including genes predicted to function in muscle structure and metabolism (Schroder *et al*. 2012).

Osteochondrosis

Osteochondrosis is a dysregulation of endochondral ossification of cartilage at the articular/epiphyseal complex most commonly occurring at the fetlock, hock and stifle joints (Jeffcott 1996). As a result, the cartilage becomes thickened and/or is retained, interfering with the normal function of the joint (Jeffcott 1996). The prevalence of OC varies by breed with relatively low incidence (~7%) in Thoroughbreds (Kane *et al*. 2003) and moderate frequency (~30%) in Danish Warmbloods (van Grevenhof *et al*. 2009), and with estimates of as many as 50% of Standardbred and German coldblooded horses (Wittwer *et al*. 2006; Lykkjen *et al*. 2012) being affected. Although it can be quite common, the heritability of OC is low to moderate (reviewed in Distl 2013; Naccache *et al*. 2018; McCoy *et al*. 2019), with a variety of environmental factors identified as having a significant role in its occurrence (Lepeule *et al*. 2009, 2013; Vander Heyden *et al*. 2013). Incidence has been positively correlated with size (Stock *et al*. 2005) and *LCORL*, itself associated with the size of a horse (described above), has been noted as a risk factor for OC and associated with incidence in GWAS (Teyssedre *et al*. 2012; Orr *et al*. 2013; Naccache *et al*. 2018). In Thoroughbreds, a QTL also on ECA3, although over 20 Mb distant from *LCORL*, was estimated to explain over 30% of the genetic variation (Corbin *et al*. 2012). Several research groups have been working to identify genetic risk factors for OC. The complexity as well as hypothesized population-specific risk factors are evident in the many loci associated using genome-wide SNP assays, many of which do not overlap between populations (Dierks *et al*. 2007, 2010; Wittwer *et al*. 2007; Lampe *et al*. 2009a,b; Lykkjen *et al*. 2010; Corbin *et al*. 2012; Teyssedre *et al*. 2012; Orr *et al*. 2013; McCoy *et al*. 2016; Lewczuk *et al*. 2017; Table 4).

Functional studies show differential expression of the *MMP-13* gene, encoding for a matrix metallopeptidase, in the cartilage of horses with OC compared with controls (Mirams *et al*. 2009; Riddick *et al*. 2012), consistent with hypothesized dysfunction in cartilage maturation and endochondral ossification. Candidate gene expression studies also suggest that chondrocyte maturation and catabolism are altered through dysregulated Wnt signaling (Kinsley *et al*. 2015). Finally, Mirams *et al*. (2016) used subtractive hybridization of the transcriptome from cartilage of affected and unaffected foals resulting in a hypothesized etiology that involves cartilage retention in subchondral bone. In addition to protein coding transcripts, differentially expressed miRNAs have been identified and may play a role in alteration of gene expression associated with OC (Desjardin *et al*. 2014).

Equine metabolic syndrome and laminitis

Horses affected by equine metabolic syndrome (EMS) present with insulin resistance and obesity and/or regional adiposity; additionally, hypertriglyceridemia, elevated leptin and hypertension may occur (Frank *et al*. 2010). Horses with EMS have an increased risk of laminitis or the disruption of the attachment between the distal phalanx (coffin bone) and inner hoof wall, leading to the coffin bone being rotated downward into the sole of the hoof. The incidence of EMS and associated laminitis is greatest in ponies and obese horses (Treiber *et al*. 2006; Bailey *et al*. 2008). Not all cases of laminitis, however, are attributed to EMS as it can occur in conjunction with other wellbeing issues such as colic, abdominal or reproductive infection, or excessive concussion on hard surfaces (Hood 1999).

SNP-based heritability estimates for traits associated with EMS (e.g. circulating insulin, glucose, ACTH, leptin) have been reported to be quite high (Norton *et al*. 2019b). Genetic factors contributing to risk are also supported by varied rates of prevalence among breeds and variation in insulin responsiveness by breed (Treiber *et al*. 2006; Bailey *et al*. 2008; Bamford *et al*. 2014). In a study of pedigreed ponies, Treiber *et al*. (2006) proposed that one or a few major dominant genes contribute to predisposition to laminitis.

Changes in gene expression related to the onset of laminitis have been studied as a possible means of both understanding the progression of the condition and identifying biomarkers to detect a possible bout prior to the onset of clinical symptoms. With a hypothesized inflammatory component, Tadros *et al*. (2013) found that circulating cytokine expression of IL-1B, IL-8 and IL-10 was elevated several hours prior to the detectable onset of laminitis. A similar inflammatory response has been identified in the laminar tissue itself in horses induced to develop laminitis compared with healthy controls (Belknap *et al*. 2007; Loftus *et al*. 2007; Leise *et al*. 2010). These studies show a systemic inflammatory response associated with laminitis. However, whereas these studies help identify pathways involved in disease progression, they fail to answer questions of genetic risk.

Toward the goal of understanding genetic risk factors for disease, Lewis *et al*. (2017) identified a single candidate gene, *FAM174A*, in Arabian horses associated with laminitis and correlated with an increased insulin-to-glucose ratio. However, a more complete understanding of the genetic risk factors for EMS remains elusive. Finally, the influence of the microbiome on the development of EMS and the associated laminitis is a recent area of study. Characterization of the hindgut microbiome revealed differences in microbial composition in horses with chronic (Steelman *et al*. 2012) and induced (Milinovich *et al*. 2006) laminitis, as well as in horses with EMS (Elzinga *et al*. 2016) compared with healthy controls. The composition of the microbiome itself has been suggested to be heritable (Blekhman *et al*. 2015;

Table 4 Complex equine diseases and traits with ongoing genetic studies.

Disease/trait (reference)	Breed	Type of genetic study	Genomic region(s) identified
Atrial fibrillation (Physick-Sheard *et al*. 2014; Kraus *et al*. 2017)	Standardbred	Heritability only	Probably polygenic; no region identified to date
Body size (e.g. Makvandi-Nejad *et al*. 2012)	Multiple	GWAS	Loci on ECA3, 6, 9 and 11
Bone fracture (Blott & Vaudin 2013; Blott *et al*. 2014)	Thoroughbred	GWAS	ECA18
Brachygnathia (Signer-Hasler *et al*. 2014)	Franches-Montagnes	GWAS	ECA13
Chronic progressive lymphedema (De Keyser *et al*. 2014),[1]	Draft breeds	Candidate gene approach	Continuing to pursue sequencing *ELN*
Common variable immunodeficiency[2]	Various	Epigenetic investigation	RNA-seq and Methyl-Seq of *E2A* and *PAX5*
Corneal stromal loss (Lassaline-Utter *et al*. 2014; Alberi *et al*. 2018)	Friesian	Heritability only/ candidate gene	Likely heritable; excluded *BGN* variants
Cribbing (crib-biting) (Hemmann *et al*. 2014)	Multiple	Candidate gene	Excluded subset of stereotypic genes
Cryptorchidism[1]	Icelandic	Heritability only	Likely to be heritable
Degenerative joint disease (Welsh *et al*. 2013)	Thoroughbred	Heritability only	Small to moderate heritability identified
Guttural pouch tympany (Metzger *et al*. 2012)	Arabians and German Warmbloods	GWAS	ECA15 (Arabians) and ECA3 (German Warmbloods)
Insect bite hypersensitivity (Schurink *et al*. 2012; Velie *et al*. 2016)	Icelandic, Shetland and Exmoor	GWAS	ECA7, 9, 10 and 17ECA8 (Exmoor)
Recurrent laryngeal neuropathy (Dupuis *et al*. 2011; Boyko *et al*. 2014)	Thoroughbred, Warmblood, Trotter and Draft	GWAS	ECA3 (height locus; Thoroughbred),ECA21 and 31 (Multiple breeds)
Metabolic syndrome (Lewis *et al*. 2017; Norton *et al*. 2019b,a)[1]	Welsh pony, Arabian, Morgan	GWAS	ECA6 (Welsh pony),ECA14 (Arabian),multiple regions (Morgan)
Navicular disease (Diesterbeck *et al*. 2007; Lopes *et al*. 2009, 2010)	Warmbloods	GWAS	ECA2 and ECA10
Neuroaxonal dystrophy/equine degenerative myeloencephalopathy (Finno *et al*. 2013, 2014)	Quarter Horse	GWAS	ECA8 region exclusion, exclusion of *TTPA* candidate gene
Osteochondrosis/osteochondrosis dissecans (Dierks *et al*. 2007; Lampe *et al*. 2009a,b,c; Sevane *et al*. 2016, 2017)	Warmbloods, Trotters, Standardbreds, Spanish Purebred	GWAS	ECA2, 4, 5, 16, 18 (Warmblood),ECA10, 14, 21 (Standardbred),candidate gene analysis (Spanish Purebred)
Polysaccharide storage myopathy, type II[1]	Quarter Horses	GWAS	ECA18
Recurrent airway obstruction (Swinburne *et al*. 2009; Schnider *et al*. 2017; Mason *et al*. 2018)	Warmblood	GWAS	ECA 11, 13, 15 (Warmblood)
Recurrent exertional rhabdomyolysis (Fritz *et al*. 2012)	Thoroughbred, Standardbred	GWAS	ECA11, 16, 30 (Thoroughbred),ECA10, 11 (Standardbred)
Recurrent uveitis (Kulbrock *et al*. 2013; Fritz *et al*. 2014)	Appaloosa, German Warmblood	Candidate gene/ GWAS	ECA1, 20 (Appaloosa),ECA18, 20 (German Warmblood)
Sarcoid (Christen *et al*. 2013; Staiger *et al*. 2016b)	Franches-Montagnes, Quarter Horse, Thoroughbred	GWAS	ECA20, 22 (QH, TB)
Stallion subfertility owing to impaired acrosome reaction (Raudsepp *et al*. 2012)	Thoroughbred	Susceptibility gene	ECA13: *FKBP6*
Stallion fertility (Schrimpf *et al*. 2015)	Hanoverian	GWAS	ECA13: *FKBP6*
Stallion fertility (Schrimpf *et al*. 2016)	19 European breeds	GWAS	High-impact variants in *CFTR* (ECA4), *OVGP1* (ECA5), *FBXO43* (ECA9), *TSSK6* (ECA21), *PKD1* (ECA13), *FOXP1* (ECA16), *TCP11* (ECA20), *SPATA31E1* (n/a), *NOTCH1* (ECA25)
Swayback (lordosis) (Cook *et al*. 2010)	Saddlebred	GWAS	ECA20

ECA, *Equus caballus* chromosome; GWA, genome-wide association.
[1]Abstract presented at the Dorothy Russell Havemeyer Foundation International Equine Genome Mapping Workshop.
[2]Abstract presented at the Plant and Animal Genome Conference.

Goodrich *et al.* 2016), associated with genes involved in metabolism and immunity; these data connect host genetics to yet another means by which risk for EMS and/or laminitis may be amplified.

Equine asthma or RAO

Equine asthma, also known as RAO or heaves, is a chronic disease of the lower airway, particularly problematic in

environments where air flow is limited and in which bedding or hay has high levels of dust or other respiratory irritants (reviewed in Woods *et al*. 1993; Ramseyer *et al*. 2007; Pirie 2014). The prevalence of RAO is estimated to be 14% in a sample of British horses (Hotchkiss *et al*. 2007) and 10% in Swiss Warmbloods (Ramseyer *et al*. 2007). Often compared with asthma in humans, affected individuals have difficulty breathing, neutrophil and mucus accumulation in the airway, cholinergic bronchospasm and coughing, and are especially sensitive to inhaled allergens (Robinson *et al*. 1996; Gerber *et al*. 2004). Whereas symptoms can be mediated by removing the horse from the problematic environment (Vandenput *et al*. 1998a, b), observations of increased risk in foals with affected parents, or foals of particular families (Marti *et al*. 1991; Ramseyer *et al*. 2007), have suggested a heritable component to its etiology.

Before genomic tools, a hypothesis derived from the clinical accumulation of mucus in RAO-affected horses led to the investigation of mucin gene variation as a possible risk factor. As a result, the equine ortholog of *MUC5AC* was identified as being upregulated in affected horses (Gerber *et al*. 2003). A candidate gene study also found an isoform of *MYH11* to be overexpressed in affected horses (Boivin *et al*. 2014). The origin of the *MYH11* isoform identified in Boivin *et al*. (2014) has been associated with alternative regulatory mechanisms (Issouf *et al*. 2018), providing an example of how variable genome regulation contributes to important phenotypes associated with animal health and wellbeing.

In perhaps the most highly studied populations of horses, and prior to the genome assembly, the candidate gene *IL4R* was investigated in two half-sibling families of Warmbloods using microsatellite genotyping (Jost *et al*. 2007). In one family, a significant association with RAO was found with a haplotype in ECA13 near the cytokine *IL4R*, and a recessive mode of inheritance was suggested. However, this association was not consistent in the other family, where an association with RAO was identified in ECA15 with a predicted dominant mode of inheritance. In a GWAS of these families using microsatellites, QTL were identified on ECA13, and whereas no association was found with the positional candidate gene *ITGAX*, *IL4R* was noted to be proximal to the hit (Swinburne *et al*. 2009). In Shakhsi-Niaei *et al*. (2012) the same research group used SNP50 to fine-map the region, which resulted in the identification of a signal on ECA13 in both the family from which the original QTL was identified and a group of unrelated horses; however, the result was not statistically significant and no clear casual mutation or genes were identified. Chromosome 13 was again the strongest region of association in a GWAS repeated on these horses with the high-density Affymetrix SNP array after it became available (Schnider *et al*. 2017); the authors note positional candidate gene, *TXNDC11*, from those analyses.

Sequencing of *IL4R* and expression from bronchoalveolar lavage fluid identified increased expression in horses from which the association was derived but not in other horse families (Klukowska-Rotzler *et al*. 2012), further supporting its hypothesized role in RAO in this population. Additional follow-up studies included quantitative PCR of candidate genes (Lanz *et al*. 2013) and RNA-seq (Pacholewska *et al*. 2015) of peripheral blood mononuclear cells (PBMCs) derived from RAO-affected and control horses that were stimulated with irritants such as hay dust or lipopolysaccharide. The role of *IL4R* in the population in which the QTL was identified was supported whereas other data continued to support different mechanisms of disease between the two families (Lanz *et al*. 2013). RNA-seq of horses from the sample families found *CXCL13* to be upregulated along with cell cycle regulatory transcripts such as *CDC20*, and genes involved in immune function and development (Pacholewska *et al*. 2015). Finally, Mason *et al*. (2018) conducted expression QTL studies of experimentally treated PBMCs from horses with prior RAO and healthy controls; whereas the studies supported prior SNP associations in these families, the identification of functional risk markers remains elusive.

Lastly, considering the possible impact of variations other than nucleotide substitutions, genomic copy number variants were analyzed using a tiling array to compare variants of RAO with control horses (Ghosh *et al*. 2014). Over 700 CNVs were identified across samples with the RAO horses found to have, on average, more CNVs than the controls. Although no significant associations were identified, *NME7*, involved in ciliary function of the lung, had a suggestive ($P = 0.06$) association with the RAO phenotype, warranting further investigation (Ghosh *et al*. 2016).

Reproduction

Compared with other complex traits, the genomics of equine reproduction has been given relatively less attention, even though reproductive performance is of high economic importance for purebred horses and vital for survival in feral populations (Raudsepp *et al*. 2013; Metzger *et al*. 2015). Only a few candidate loci or genomic regions have been associated with various fertility parameters and phenotypes (Raudsepp *et al*. 2013). Among these, the *FKBP6* gene is of particular interest because of contrasting associations: in Thoroughbreds it is associated with subfertility owing to impaired acrosome reaction (Raudsepp *et al*. 2012), but in Hanoverians it is associated with improved conception rates (Schrimpf *et al*. 2015). Like for other complex traits, genome-wide SNP- or WGS-based analyses are expected to also make a difference in fertility research. A good example is a recent whole-genome screening that revealed high-impact variants in nine putative male fertility genes (Schrimpf *et al*. 2016; Table 4).

On a different note, breeding animals are typically selected on the basis of their pedigrees, athletic performance and appearance, but not for their reproductive potential. This suggests that there are no strong signatures of selection for reproductive performance. Surprisingly, this is not true as shown by a whole-genome analysis of runs of homozygosity in six diverse breeds, including commercial breeds such as Hanoverian and Thoroughbred, and native breeds such as Sorraia (Metzger *et al.* 2015). The findings suggest a significant artificial as well as natural positive selection on reproduction performance in all types of horse populations.

Horse as a large animal model for humans

Unlike rats or mice, horses are large and expensive animals with long generation intervals, and therefore are not typical model species for studying human disease and physiology. However, some equine conditions, such as obesity, respiratory disease, orthopedic disease, equine recurrent uveitis and certain cancers translate into similar human conditions better than those from classical model species. For example, because of unique similarities between human and horse insulin resistance response to overfeeding, EMS has the potential to serve as a model for human obesity (Frank *et al.* 2010; Jacob *et al.* 2018). Likewise, naturally occurring equine asthma is recognized as a model for some forms of asthma in humans (Bullone & Lavoie 2017; Bond *et al.* 2018). As an athletic species, the horse is considered as an important large animal model for cardiovascular disease (Tsang *et al.* 2016) and musculoskeletal disorders, including osteoarthritis (McCoy 2015), as well as a model for articular cartilage repair and regeneration studies (Dias *et al.* 2018). The horse has also been proposed as a potential model for immune response for infections, such as acute synovitis and septic arthritis (Ludwig *et al.* 2016), and autoimmune disorders like recurrent uveitis (Witkowski *et al.* 2016). Furthermore, studies of melanoma in gray horses are expected to help dissect the molecular mechanisms underlying melanoma as well as vitiligo in humans (Rosengren Pielberg *et al.* 2008; van der Weyden *et al.* 2016).

Functional annotation of the equine genome

Despite the recent and rapid successes in identifying causative mutations for these mostly simple inherited traits, the genetic investigation of complex traits has not been as straightforward. Although there is strong evidence for the heritability of many complex diseases and traits (described above), despite a significant financial investment by the equine industry in phenotyping and genotyping large numbers of horses, functional genetic mutations have not been discovered for most complex traits examined (Table 4).

Finding genetic associations of chromosomal segments to a specific phenotype without identifying functional or causative mutations within protein-coding genes is not a problem unique to the horse. The idea that much of the non-coding genome is 'junk DNA' (Ohno 1972) and uninteresting to consider further has been reconsidered after overwhelming evidence from the human and murine Encyclopedia of DNA Elements (ENCODE) projects, demonstrating that these regions are functionally important (Consortium *et al.* 2007; Yue *et al.* 2014). As many as 93% of the associated and potentially causative variants from human GWAS publications are outside of annotated protein-coding regions (Hindorff *et al.* 2009). It is unlikely that these associations are aberrant, but rather they are identifying genomic regions involved in processes other than protein coding, such as regulation of gene expression (Maurano *et al.* 2012). In addition to the recent discovery that the majority of the genome is transcribed (Consortium *et al.* 2007), it has also been shown that 92–94% of human protein-coding genes express multiple mRNA variants or isoforms (Wang *et al.* 2008). Many of the variants responsible for mapped equine diseases and traits in Table 3, in addition to others, may be regulatory, non-coding mutations.

The need for defining the tissue-specific gene expression and regulation (i.e. functional annotation) across domestic animal species was acknowledged by the establishment of the International FAANG project (Andersson *et al.* 2015). The ultimate goal of this initiative is to provide high-quality functional annotation of animal genomes in a coordinated effort that facilitates data sharing and analysis, while establishing standards for assay quality and continuity for metadata analysis. In an effort to establish a more direct connection between genome function and phenotype, the FAANG consortium has focused on initially assaying tissues from one to two individual animals representing a breed with minimal genetic diversity within a species (Andersson *et al.* 2015). The resulting data from this annotation across species will provide power to associate phenotypes with functional data, making it possible to generate and test hypotheses regarding the functional mechanisms underlying associations (Tuggle *et al.* 2016; Giuffra *et al.* 2019).

As demonstrated by the human and murine ENCODE projects, five types of assays uncover the majority of variation in tissue-specific gene expression and epigenetic modifications:

1 *Expression*—RNA-seq identifies expression levels of primarily protein-coding genes.
2 *RNA expression*—small RNA-seq identifies expression of miRNAs, non-coding RNAs that function primarily in RNA silencing.
3 *Histone modifications*—ChIP-seq identifies genome-wide patterns of histone modifications using antibodies against the modified histone proteins. These prioritized marks have been standardized across species and those focused upon for the FAANG initiative and the genomic features they identify include H3K4me1 (enhancers and

Table 5 Biobank tissues collected from two Thoroughbred mares.

Musculoskeletal system:	Nervous system:	Digestive system:
Cartilage — Fetlock	*Cerebellum Vermis*	*Cecum*
Cartilage — Stifle	*Cerebellum Lateral Hemisphere*	*Duodenum*
Coronary Band	*Cornea*	Epiglottis
Deep Digital Flexor	Corpus Callosum	Esophagus
Gluteal Muscle	*C1 Spinal Cord*	*Ileum*
Lamina	C6 Spinal Cord	*Jejunum*
Long Bone Marrow	*Dorsal Root Ganglia*	Left Dorsal Colon
Longissimus Dorsi	Dura Mater	Left Ventral Colon
Metacarpal Bone Diaphysis	*Frontal Cortex*	Right Dorsal Colon
Rib Bone Marrow	*Hypothalamus*	Right Ventral Colon
Sacrocaudalis Muscle	L1 Spinal Cord	Small Colon
Sesamoid Bone	*L6 Spinal Cord*	Stomach
Superficial Digital Flexor	*Occipital Cortex*	Tongue
Suspensory Ligament	**Parietal Cortex**	Abdominal/thoracic organs:
Cardiovascular system:	*Pituitary*	*Adrenal Cortex*
Aortic Valve	Pons	Adrenal Medulla
Heart Left Atrium	*Retina*	*Kidney Cortex*
Heart Left Ventricle	Sciatic Nerve	*Kidney Medulla*
Heart Right Atrium	*Temporal Cortex*	*Larynx*
Heart Right Ventricle	Thalamus	**Liver**
Left Lung	*T8 Spinal Cord*	Lymph Node
Mitral Valve	Cell types:	Pancreas
Pulmonic Valve	*Fibroblasts (culture)*	*Spleen*
Trachea	*Keratinocytes (culture)*	Thyroid
Tricuspid Valve	*PBMCs*	Integumentary system:
Urogenital System:	Body fluids:	*Dorsum Skin (over back)*
Cervix	Plasma	Gluteal Adipose
Mammary Gland	Serum	**Loin Adipose**
Ovary	Cerebrospinal fluid	Neck Skin
Oviduct	Synovial fluid	
Urinary Bladder		
Uterus		
Vagina		

Updated from Burns *et al*. (2018). Prioritized tissues for study are in bold. Additional tissues from which RNA-seq data have been collected as funded by outside collaborators are shown in italics.

distal regulatory elements), H3K4me3 (active promoters and enhancers), H3K27me3 (gene silencing) and H3K27ac (active regulatory elements; Lee *et al*. 2014).

4 *Chromatin accessibility*—the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) identifies regions of open chromatin.

5 *DNA methylation*—reduced representation bisulfite sequencing identifies DNA methylation across the genome.

In line with the priorities of the FAANG initiative, in 2016, an equine biobank was created based on the sampling and preservation of 86 tissues, two cell lines and fluids from two Thoroughbred mares (Table 5; Burns *et al*. 2018). Tissues were flash frozen, preserved for histopathology, fixed for chromatin-immunoprecipitation, and in 16 tissues, nuclei isolation was conducted. This biobank is available for all researchers to access for assays appropriate to the goals of the FAANG initiative. Notably, extensive ante- and post-mortem evaluation by veterinarians was conducted on these two horses to provide the highest standard of a true 'reference' database for researchers to utilize. In addition, both mares represented in the biobank had normal karyotypes. Utilizing the strict standard of phenotyping allows for both association of phenotype with genomic data and standardization of future sampling efforts, enhancing the utility of the data generated (Burns *et al*. 2018).

Eight tissues were prioritized in the initial equine annotation efforts (Table 5) owing to their cross-species application (e.g. skeletal muscle, liver, lung and ovary) as well as importance to the horse (laminae). Additionally, as an international collaborative initiative, 24 individuals representing 16 research institutions in 10 countries voluntarily participated to support RNA-seq of additional tissues by providing funding for transcriptome characterization of the tissue(s) most relevant to their own research; RNA-seq datasets from these tissues are publicly available (EMBL: http://www.ebi.ac.uk/embl/). Seven laboratories also conducted additional assays, such as karyotype analyses, centromere mapping of fibroblast cells, reduced representation bisulfite sequencing of the eight priority tissues, fibroblast functional assays, functional assays on tissues of the suspensory apparatus and further phenotyping through the sequencing of microbiome samples.

ChIP assays for four histone modification marks were recently completed on the eight prioritized tissues. At this time, ChIP assays for the major insulator-binding protein in vertebrates, CCCTC-binding factor and ATAC-seq experiments are underway. All datasets will then be fully integrated and correlated with gene expression data and made publicly available as an equine-specific tissue atlas to the entire equine research community.

Future directions

With the sheer volume of sequencing efforts currently underway in the horse, in addition to the FAANG efforts to define regulatory regions of the genome, the genetic contributions to complex traits can be discovered. Many of these more complex diseases (Table 4) probably have strong environmental contributions to the overall phenotype. Therefore, educating veterinarians and horse owners on the proper use of these genetic 'risk factors' in breeding management will be essential to advance the health of horses.

Concluding remarks and future perspectives

In this review, we demonstrate that the 10 years of post-genome era in equine genomics have been unparalleled and decorated with outstanding achievements in almost all conceivable directions. These include improved characterization of the structure and function of the horse genome, delineating the genetic makeup of breeds and populations, deciphering the evolutionary ancestry of horses and continuing search for molecular causes of Mendelian and complex traits and diseases. The central pillar of support for this success is definitely the high-quality horse reference genome combined with unprecedented advances in genomics technologies and global collaborations between researchers of diverse disciplines, clinicians, breeders and horse owners.

Whereas similar achievements characterize the post-genome era of all main domestic species, it must be emphasized that the collection of genome-level sequence data from hundreds of ancient horses from the past few thousand years is unparalleled and not available for any other domestic or non-primate species. This unique resource allows researchers to track the evolutionary past of any genomic features, particularly the sequence variants that are associated with equine traits of importance, such as performance, coat color, disease and adaptations. This also demonstrates that the history of domestic animals cannot be fully understood without ancient genomic data.

The enhanced knowledge about the horse genome, biology and populations also highlights the areas that require critical improvement. Among these, high expectations are put on the equine FAANG initiative in order to identify important functional elements in the horse genome, particularly those underlying simple and complex traits. Also, the growing number of available WGSs of individual horses of diverse breeds and phenotypes allows for a comprehensive catalog of common and rare variants in the horse genome. This in turn is the foundation for equine precision medicine, which should identify novel genetic mutations in a small number of individuals and connect the variation with function.

Acknowledgements

Funding for TR was provided by USDA-NIFA awards 2012-67015-19632 and 2019-67015-29322, and for CJF by NIH K01OD015134 and LRP L40 TR001136.

Conflict of interests

The authors declare no conflict of interests.

References

Achilli A., Olivieri A., Soares P. *et al.* (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2449–54.

Alberi C., Hisey E., Lassaline M., Atilano A., Kalbfleisch T., Stoppini R., Hermans H., Back W., Mienaltowski M.J. & Bellone R.R. (2018) Ruling out BGN variants as simple X-linked causative mutations for bilateral corneal stromal loss in Friesian horses. *Animal Genetics* **49**, 656–7.

Almarzook S., Reissmann M., Arends D. & Brockmann G.A. (2017) Genetic diversity of Syrian Arabian horses. *Animal Genetics* **48**, 486–9.

Andersson L.S., Larhammar M., Memic F. *et al.* (2012) Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature* **488**, 642–6.

Andersson L., Archibald A.L., Bottema C.D. *et al.* (2015) Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biology* **16**, 57.

Árnason T. (2001) Genetic evaluation of Swedish standard-bred trotters for racing performance traits and racing status. *Journal of Animal Breeding and Genetics* **116**, 387–98.

Avila F., Mickelson J.R., Schaefer R.J. & McCue M.E. (2018) Genome-wide signatures of selection reveal genes associated with performance in American quarter horse subpopulations. *Frontiers in Genetics* **9**, 249.

Bailey E. & Brooks S.A. (2013) *Horse Genetics*, 2nd edn. CABI, Oxford.

Bailey S.R., Habershon-Butcher J.L., Ransom K.J., Elliott J. & Menzies-Gow N.J. (2008) Hypertension and insulin resistance in a mixed-breed population of ponies predisposed to laminitis. *American Journal of Veterinary Research* **69**, 122–9.

Balmer P., Bauer A., Pujar S., McGarvey K.M., Welle M., Galichet A., Muller E.J., Pruitt K.D., Leeb T. & Jagannathan V. (2017) A curated catalog of canine and equine keratin genes. *PLoS ONE* **12**, e0180359.

Bamford N.J., Potter S.J., Harris P.A. & Bailey S.R. (2014) Breed differences in insulin sensitivity and insulinemic responses to oral

glucose in horses and ponies of moderate body condition score. *Domestic Animal Endocrinology* **47**, 101–7.

de Barros Damgaard P., Martiniano R., Kamm J. *et al*. (2018) The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**. https://doi.org/10.1126/science.aar7711

Bauer A., Hiemesch T., Jagannathan V. *et al*. (2017) A nonsense variant in the ST14 gene in Akhal-Teke horses with naked foal syndrome. *G3 (Bethesda)* **7**, 1315–21.

Beeson S.K., Schaefer R.J., Mason V.C. & McCue M.E. (2019) Robust remapping of equine SNP array coordinates to EquCab3.0. *Animal Genetics* **50**, 114–5.

Belknap J.K., Giguere S., Pettigrew A., Cochran A.M., Van Eps A.W. & Pollitt C.C. (2007) Lamellar pro-inflammatory cytokine expression patterns in laminitis at the developmental stage and at the onset of lameness: innate vs. adaptive immune response. *Equine Veterinary Journal* **39**, 42–7.

Bellone R.R., Holl H., Setaluri V. *et al*. (2013) Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLoS ONE* **8**, e78280.

Bellone R.R., Liu J., Petersen J.L. *et al*. (2017) A missense mutation in damage-specific DNA binding protein 2 is a genetic risk factor for limbal squamous cell carcinoma in horses. *International Journal of Cancer* **141**, 342–53.

Beltran N.A.R., Meira C.T., de Oliveira H.N., Pereira G.L., Silva J.A.I.V., da Mota M.D.S. & Curi R.A. (2015) Prospection of genomic regions divergently selected in cutting line of Quarter Horses in relation to racing line. *Livestock Science* **174**, 1–9.

Binns M.M., Boehler D.A. & Lambert D.H. (2010) Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Animal Genetics* **41** (Suppl 2), 154–8.

Blekhman R., Goodrich J.K., Huang K. *et al*. (2015) Host genetic variation impacts microbiome composition across human body sites. *Genome Biology* **16**, 191.

Blott S.C. & Vaudin M. (2013) Genomic selection against fracture risk in the Thoroughbred horse. *BMC Genomics* **15**, 1–8.

Blott S.C., Swinburne J.E., Sibbons C., Fox-Clipsham L.Y., Helwegen M., Hillyer L., Parkin T.D., Newton J.R. & Vaudin M. (2014) A genome-wide association study demonstrates significant genetic variation for fracture risk in Thoroughbred racehorses. *BMC Genomics* **15**, 147.

Boeskorov G.G., Potapova O.R., Protopopov A.V., Plotnikov V.V., Maschenko E.N., Shchelchkova M.V., Petrova E.A., Kowalczyk R., van der Plicht J. & Tikhonov A.N. (2018) A study of a frozen mummy of a wild horse from the Holocene of Yakutia, East Siberia, Russia. *Mammal Research* **63**, 1–8.

Boivin R., Vargas A., Lefebvre-Lavoie J., Lauzon A.M. & Lavoie J.P. (2014) Inhaled corticosteroids modulate the (+)insert smooth muscle myosin heavy chain in the equine asthmatic airways. *Thorax* **69**, 1113–9.

Bond S., Leguillette R., Richard E.A., Couetil L., Lavoie J.P., Martin J.G. & Pirie R.S. (2018) Equine asthma: integrative biologic relevance of a recently proposed nomenclature. *Journal of Veterinary Internal Medicine* **32**, 2088–98.

Bordbari M.H., Penedo M.C.T., Aleman M., Valberg S.J., Mickelson J. & Finno C.J. (2017) Deletion of 2.7 kb near HOXD3 in an Arabian horse with occipitoatlantoaxial malformation. *Animal Genetics* **48**, 287–94.

Boyko A.R., Brooks S.A., Behan-Braman A. *et al*. (2014) Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC Genomics* **15**, 259.

Braam A., Nasholm A., Roepstorff L. & Philipsson J. (2011) Genetic variation in durability of Swedish Warmblood horses using competition results. *Livestock Science* **142**, 181–7.

Brakenhoff J.E., Holcombe S.J., Hauptman J.G., Smith H.K., Nickels F.A. & Caron J.P. (2006) The prevalence of laryngeal disease in a large population of competition draft horses. *Veterinary Surgery* **35**, 579–83.

Brandariz-Fontes C., Leonard J.A., Vega-Pla J.L., Backstrom N., Lindgren G., Lippold S. & Rico C. (2013) Y-chromosome analysis in retuertas horses. *PLoS ONE* **8**, e64985.

Brard S. & Ricard A. (2015) Genome-wide association study for jumping performances in French sport horses. *Animal Genetics* **46**, 78–81.

Brooks S.A., Gabreski N., Miller D., Brisbin A., Brown H.E., Streeter C., Mezey J., Cook D. & Antczak D.F. (2010) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genetics* **6**, e1000909.

Brosnahan M.M., Brooks S.A. & Antczak D.F. (2010) Equine clinical genomics: a clinician's primer. *Equine Veterinary Journal* **42**, 658–70.

Bullone M. & Lavoie J.P. (2017) The contribution of oxidative stress and inflamm-aging in human and equine asthma. *International Journal of Molecular Sciences* **18**, 1–23.

Burns E.N., Bordbari M.H., Mienaltowski M.J. *et al*. (2018) Generation of an equine biobank to be used for Functional Annotation of Animal Genomes project. *Animal Genetics* **49**, 564–70.

Cerutti F., Gamba R., Mazzagatti A., Piras F.M., Cappelletti E., Belloni E., Nergadze S.G., Raimondi E. & Giulotto E. (2016) The major horse satellite DNA family is associated with centromere competence. *Molecular Cytogenetics* **9**, 35.

Chowdhary B.P. (2013) *Equine Genomics*. Wiley-Blackwell, Ames.

Chowdhary B.P. & Bailey E. (2003) Equine genomics: galloping to new frontiers. *Cytogenetic and Genome Research* **102**, 184–8.

Chowdhary B.P. & Raudsepp T. (2006) The horse genome. *Genome Dynamics* **2**, 97–110.

Chowdhary B.P. & Raudsepp T. (2008) The horse genome derby: racing from map to whole genome sequence. *Chromosome Research* **16**, 109–27.

Chowdhary B.P., Paria N. & Raudsepp T. (2008) Potential applications of equine genomics in dissecting diseases and fertility. *Animal Reproduction Science* **107**, 208–18.

Christen G., Gerber V., Dolf G., Burger D. & Koch C. (2013) Inheritance of equine sarcoid disease in Franches-Montagnes horses. *The Veterinary Journal* **199**, 68–71.

Coleman S.J., Zeng Z., Wang K., Luo S., Khrebtukova I., Mienaltowski M.J., Schroth G.P., Liu J. & MacLeod J.N. (2010) Structural annotation of equine protein-coding genes determined by mRNA sequencing. *Animal Genetics* **41** (Suppl 2), 121–30.

Coleman S.J., Mienaltowski M.J. & MacLeod J.N. (2013a) Functional genomics. In: *Equine Genomics* (Ed. by B.P. Chowdhary), pp. 125–41. John Wiley & Sons, Inc., Blackwell Publishing, Ames, IA.

Coleman S.J., Zeng Z., Hestand M.S., Liu J. & Macleod J.N. (2013b) Analysis of unannotated equine transcripts identified by mRNA sequencing. *PLoS ONE* **8**, e70125.

Consortium E.P., Birney E., Stamatoyannopoulos J.A. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816.

Cook D., Gallagher P.C. & Bailey E. (2010) Genetics of swayback in American Saddlebred horses. *Animal Genetics* **41** (Suppl 2), 64–71.

Corbin L.J., Blott S.C., Swinburne J.E. *et al.* (2012) A genome-wide association study of osteochondritis dissecans in the Thoroughbred. *Mammalian Genome* **23**, 294–303.

De Keyser K., Janssens S., Peeters L.M., Foque N., Gasthuys F., Oosterlinck M. & Buys N. (2014) Genetic parameters for chronic progressive lymphedema in Belgian Draught Horses. *Journal of Animal Breeding and Genetics* **131**, 522–8.

Der Sarkissian C., Ermini L., Schubert M. *et al.* (2015) Evolutionary genomics and conservation of the endangered Przewalski's horse. *Current Biology* **25**, 2577–83.

Desjardin C., Vaiman A., Mata X. *et al.* (2014) Next-generation sequencing identifies equine cartilage and subchondral bone miRNAs and suggests their involvement in osteochondrosis physiopathology. *BMC Genomics* **15**, 798.

Dias I.R., Viegas C.A. & Carvalho P.P. (2018) Large animal models for osteochondral regeneration. *Advances in Experimental Medicine and Biology* **1059**, 441–501.

Dierks C., Lohring K., Lampe V., Wittwer C., Drogemuller C. & Distl O. (2007) Genome-wide search for markers associated with osteochondrosis in Hanoverian warmblood horses. *Mammalian Genome* **18**, 739–47.

Dierks C., Komm K., Lampe V. & Distl O. (2010) Fine mapping of a quantitative trait locus for osteochondrosis on horse chromosome 2. *Animal Genetics* **41**, 87–90.

Diesterbeck U.S., Hertsch B. & Distl O. (2007) Genome-wide search for microsatellite markers associated with radiologic alterations in the navicular bone of Hanoverian warmblood horses. *Mammalian Genome* **18**, 373–81.

van Dijk E.L., Auger H., Jaszczyszyn Y. & Thermes C. (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* **30**, 418–26.

Distl O. (2013) The genetics of equine osteochondrosis. *Veterinary Journal* **197**, 13–8.

Doan R., Cohen N., Harrington J., Veazey K., Juras R., Cothran G., McCue M.E., Skow L. & Dindot S.V. (2012) Identification of copy number variants in horses. *Genome Research* **22**, 899–907.

Drogemuller M., Jagannathan V., Welle M.M. *et al.* (2014) Congenital hepatic fibrosis in the Franches-Montagnes horse is associated with the polycystic kidney and hepatic disease 1 (PKHD1) gene. *PLoS ONE* **9**, e110125.

Ducro B.J., Schurink A., Bastiaansen J.W. *et al.* (2015) A nonsense mutation in B3GALNT2 is concordant with hydrocephalus in Friesian horses. *BMC Genomics* **16**, 761.

Dupuis M.C., Zhang Z., Druet T., Denoix J.M., Charlier C., Lekeux P. & Georges M. (2011) Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. *Mammalian Genome* **22**, 613–20.

Eberth J.E., Graves K.T., MacLeod J.N. & Bailey E. (2018) Multiple alleles of ACAN associated with chondrodysplastic dwarfism in Miniature horses. *Animal Genetics* **49**, 413–20.

Elzinga S.E., Weese J.S. & Adams A.A. (2016) Comparison of the fecal microbiota in horses with equine metabolic syndrome and metabolically normal controls fed a similar all-forage diet. *Journal of Equine Veterinary Science* **44**, 9–16.

Fages A., Hanghoj K., Khan N. *et al.* (2019) Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419–35 e31.

Fegraeus K.J., Hirschberg I., Arnason T., Andersson L., Velie B.D., Andersson L.S. & Lindgren G. (2017) To pace or not to pace: a pilot study of four- and five-gaited Icelandic horses homozygous for the DMRT3 'Gait Keeper' mutation. *Animal Genetics* **48**, 694–7.

Felkel S., Vogl C., Rigler D. *et al.* (2018) Asian horses deepen the MSY phylogeny. *Animal Genetics* **49**, 90–3.

Felkel S., Vogl C., Rigler D. *et al.* (2019) The horse Y chromosome as an informative marker for tracing sire lines. *Scientific Reports* **9**, 6095.

Finno C.J. & Bannasch D.L. (2014) Applied equine genetics. *Equine Veterinary Journal* **46**, 538–44.

Finno C.J., Spier S.J. & Valberg S.J. (2009) Equine diseases caused by known genetic mutations. *Veterinary Journal* **179**, 336–47.

Finno C.J., Famula T., Aleman M., Higgins R.J., Madigan J.E. & Bannasch D.L. (2013) Pedigree analysis and exclusion of alpha-tocopherol transfer protein (TTPA) as a candidate gene for neuroaxonal dystrophy in the American Quarter Horse. *Journal of Veterinary Internal Medicine* **27**, 177–85.

Finno C.J., Aleman M., Higgins R.J., Madigan J.E. & Bannasch D.L. (2014) Risk of false positive genetic associations in complex traits with underlying population structure: a case study. *Veterinary Journal* **202**, 543–9.

Finno C.J., Gianino G., Perumbakkam S., Williams Z.J., Bordbari M.H., Gardner K.L., Burns E., Peng S., Durward-Akhurst S.A. & Valberg S.J. (2018) A missense mutation in MYH1 is associated with susceptibility to immune-mediated myositis in Quarter Horses. *Skeletal Muscle* **8**, 7.

Fonseca M.G., Ferraz G.D., Lage J., Pereira G.L. & Curi R.A. (2017) A genome-wide association study reveals differences in the genetic mechanism of control of the two gait patterns of the Brazilian Mangalarga Marchador breed. *Journal of Equine Veterinary Science* **53**, 64–7.

Frank N., Geor R.J., Bailey S.R., Durham A.E., Johnson P.J. & American College of Veterinary Internal M. (2010) Equine metabolic syndrome. *Journal of Veterinary Internal Medicine* **24**, 467–75.

Frischknecht M., Jagannathan V., Plattet P. *et al.* (2015) A non-synonymous HMGA2 variant decreases height in Shetland ponies and other small horses. *PLoS ONE* **10**, e0140749.

Frischknecht M., Signer-Hasler H., Leeb T., Rieder S. & Neuditschko M. (2016) Genome-wide association studies based on sequence-derived genotypes reveal new QTL associated with conformation and performance traits in the Franches-Montagnes horse breed. *Animal Genetics* **47**, 227–9.

Fritz K.L., McCue M.E., Valberg S.J., Rendahl A.K. & Mickelson J.R. (2012) Genetic mapping of recurrent exertional rhabdomyolysis in a population of North American Thoroughbreds. *Animal Genetics* **43**, 730–8.

Fritz K.L., Kaese H.J., Valberg S.J. *et al.* (2014) Genetic risk factors for insidious equine recurrent uveitis in Appaloosa horses. *Animal Genetics* **45**, 392–9.

Gaffney B. & Cunningham E.P. (1988) Estimation of genetic trend in racing performance of thoroughbred horses. *Nature* **332**, 722–4.

Gaunitz C., Fages A., Hanghoj K. *et al*. (2018) Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111–4.

Gerber V., Robinson N.E., Venta R.J., Rawson J., Jefcoat A.M. & Hotchkiss J.A. (2003) Mucin genes in horse airways: MUC5AC, but not MUC2, may play a role in recurrent airway obstruction. *Equine Veterinary Journal* **35**, 252–7.

Gerber V., Straub R., Marti E., Hauptman J., Herholz C., King M., Imhof A., Tahon L. & Robinson N.E. (2004) Endoscopic scoring of mucus quantity and quality: observer and horse variance and relationship to inflammation, mucus viscoelasticity and volume. *Equine Veterinary Journal* **36**, 576–82.

Ghosh S., Qu Z., Das P.J. *et al*. (2014) Copy number variation in the horse genome. *PLoS Genetics* **10**, e1004712.

Ghosh S., Das P.J., McQueen C.M., Gerber V., Swiderski C.E., Lavoie J.P., Chowdhary B.P. & Raudsepp T. (2016) Analysis of genomic copy number variation in equine recurrent airway obstruction (heaves). *Animal Genetics* **47**, 334–44.

Ghosh M., Sharma N., Singh A.K., Gera M., Pulicherla K.K. & Jeong D.K. (2018) Transformation of animal genomics by next-generation sequencing technologies: a decade of challenges and their impact on genetic architecture. *Critical Reviews in Biotechnology* **38**, 1157–75.

Giuffra E., Tuggle C.K. & Consortium F. (2019) Functional annotation of animal genomes (FAANG): current achievements and roadmap. *Annual Review of Animal Biosciences* **7**, 65–88.

Giulotto E., Raimondi E. & Sullivan K.F. (2017) The unique DNA sequences underlying equine centromeres. *Progress in Molecular and Subcellular Biology* **56**, 337–54.

Goodrich J.K., Davenport E.R., Beaumont M., Jackson M.A., Knight R., Ober C., Spector T.D., Bell J.T., Clark A.G. & Ley R.E. (2016) Genetic determinants of the gut microbiome in UK twins. *Cell Host & Microbe* **19**, 731–43.

Goto H., Ryder O.A., Fisher A.R., Schultz B., Kosakovsky Pond S.L., Nekrutenko A. & Makova K.D. (2011) A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. *Genome Biology and Evolution* **3**, 1096–106.

van Grevenhof E.M., Ducro B.J., van Weeren P.R., van Tartwijk J.M.F.M., van den Belt A.J. & Bijma P. (2009) Prevalence of various radiographic manifestations of osteochondrosis and their correlations between and within joints in Dutch Warmblood horses. *Equine Veterinary Journal* **41**, 11–6.

Gu J., MacHugh D.E., McGivney B.A., Park S.D., Katz L.M. & Hill E.W. (2010) Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses. *Equine Veterinary Journal Supplement* **42** (Suppl. 38), 569–75.

Gurgul A., Jasielczuk I., Semik-Gurgul E., Pawlina-Tyszko K., Stefaniuk-Szmukier M., Szmatola T., Polak G., Tomczyk-Wrona I. & Bugno-Poniewierska M. (2019) A genome-wide scan for diversifying selection signatures in selected horse breeds. *PLoS ONE* **14**, e0210751.

Han H., Wallner B., Rigler D., MacHugh D.E., Manglai D. & Hill E.W. (2019) Chinese Mongolian horses may retain early domestic male genetic lineages yet to be discovered. *Animal Genetics* **50**, 399–402.

Hemmann K., Ahonen S., Raekallio M., Vainio O. & Lohi H. (2014) Exploration of known stereotypic behaviour-related candidate genes in equine crib-biting. *Animal*, **8**, 347–53.

Hendricks B.L. (2007) *International Encyclopedia of Horse Breeds*. Norman, OK: University of Oklahoma Press.

Henkel J., Lafayette C., Brooks S.A., Martin K., Patterson-Rosa L., Cook D., Jagannathan V. & Leeb T. (2019) Whole-genome sequencing reveals a large deletion in the MITF gene in horses with white spotted coat colour and increased risk of deafness. *Animal Genetics* **50**, 172–4.

Hestand M.S., Kalbfleisch T.S., Coleman S.J., Zeng Z., Liu J., Orlando L. & MacLeod J.N. (2015) Annotation of the protein coding regions of the equine genome. *PLoS ONE* **10**, e0124375.

Higuchi R., Bowman B., Freiberger M., Ryder O.A. & Wilson A.C. (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**, 282–4.

Hill E.W., Gu J., McGivney B.A. & MacHugh D.E. (2010a) Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Animal Genetics* **41**(Suppl 2), 56–63.

Hill E.W., McGivney B.A., Gu J., Whiston R. & Machugh D.E. (2010b) A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC Genomics* **11**, 552.

Hindorff L.A., Sethupathy P., Junkins H.A., Ramos E.M., Mehta J.P., Collins F.S. & Manolio T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9362–7.

Hintz R.L. & Vanvleck L.D. (1978) Factors influencing racing performance of standardbred pacer. *Journal of Animal Science* **46**, 60–8.

Hood D.M. (1999) The pathophysiology of developmental and acute laminitis. *Veterinary Clinics of North America: Equine Practice* **15**, 321–43.

Hotchkiss J.W., Reid S.W.J. & Christley R.M. (2007) A survey of horse owners in Great Britain regarding horses in their care. Part 2: risk factors for recurrent airway obstruction. *Equine Veterinary Journal* **39**, 301–8.

Issouf M., Vargas A., Boivin R. & Lavoie J.P. (2018) SRSF6 is upregulated in asthmatic horses and involved in the MYH11 SMB expression. *Physiological Reports* **6**, e13896.

Jacob S.I., Murray K.J., Rendahl A.K., Geor R.J., Schultz N.E. & McCue M.E. (2018) Metabolic perturbations in Welsh Ponies with insulin dysregulation, obesity, and laminitis. *Journal of Veterinary Internal Medicine* **32**, 1215–33.

Jaderkvist Fegraeus K., Johansson L., Maenpaa M., Mykkanen A., Andersson L.S., Velie B.D., Andersson L., Arnason T. & Lindgren G. (2015) Different DMRT3 genotypes are best adapted for harness racing and riding in Finnhorses. *Journal of Heredity* **106**, 734–40.

Jaderkvist K., Andersson L.S., Johansson A.M., Arnason T., Mikko S., Eriksson S., Andersson L. & Lindgren G. (2014) The DMRT3 'Gait keeper' mutation affects performance of Nordic and Standardbred trotters. *Journal of Animal Science* **92**, 4279–86.

Jagannathan V., Gerber V., Rieder S., Tetens J., Thaller G., Drogemuller C. & Leeb T. (2019) Comprehensive characterization of horse genome variation by whole-genome sequencing of 88 horses. *Animal Genetics* **50**, 74–7.

Jain M., Olsen H.E., Turner D.J., Stoddart D., Bulazel K.V., Paten B., Haussler D., Willard H.F., Akeson M. & Miga K.H. (2018) Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology* **36**, 321–3.

Janečka J.E., Davis B.W., Ghosh S. *et al.* (2018) Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. *Nature Communications* **9**, 2945.

Jeffcott L.B. (1996) Osteochondrosis—an international problem for the horse industry. *Journal of Equine Veterinary Science* **16**, 32–7.

Jonsson H., Schubert M., Seguin-Orlando A. *et al.* (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 18655–60.

Jost U., Klukowska-Rotzler J., Dolf G., Swinburne J.E., Ramseyer A., Bugno M., Burger D., Blott S. & Gerber V. (2007) A region on equine chromosome 13 is linked to recurrent airway obstruction in horses. *Equine Veterinary Journal* **39**, 236–41.

Kakoi H., Kikuchi M., Tozaki T., Hirota K.I., Nagata S.I., Hobo S. & Takasu M. (2018) Distribution of Y chromosomal haplotypes in Japanese native horse populations. *Journal of Equine Science* **29**, 39–42.

Kalbfleisch T.S., Rice E.S., DePriest M.S. *et al.* (2018) Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology* **1**, 197.

Kane A.J., Park R.D., McIlwraith C.W., Rantanen N.W., Morehead J.P. & Bramlage L.R. (2003) Radiographic changes in Thoroughbred yearlings. Part 1: Prevalence at the time of the yearling sales. *Equine Veterinary Journal* **35**, 354–65.

Khaudov A.D., Duduev A.S., Kokov Z.A., Amshokov K.K., Zhekamukhov M.K., Zaitsev A.M. & Reissmann M. (2018) Genetic analysis of maternal and paternal lineages in Kabardian horses by uniparental molecular markers. *Open Veterinary Journal* **8**, 40–6.

Kinsley M.A., Semevolos S.A. & Duesterdieck-Zellmer K.F. (2015) Wnt/beta-catenin signaling of cartilage canal and osteochondral junction chondrocytes and full thickness cartilage in early equine osteochondrosis. *Journal of Orthopaedic Research* **33**, 1433–8.

Klukowska-Rotzler J., Swinburne J.E., Drogemuller C., Dolf G., Janda J., Leeb T. & Gerber V. (2012) The interleukin 4 receptor gene and its role in recurrent airway obstruction in Swiss Warmblood horses. *Animal Genetics* **43**, 450–3.

Kraus M., Physick-Sheard P.W., Brito L.F. & Schenkel F.S. (2017) Estimates of heritability of atrial fibrillation in the Standardbred racehorse. *Equine Veterinary Journal* **49**, 718–22.

Kreutzmann N., Brem G. & Wallner B. (2014) The domestic horse harbours Y-chromosomal microsatellite polymorphism only on two widely distributed male lineages. *Animal Genetics* **45**, 460.

Kulbrock M., Lehner S., Metzger J., Ohnesorge B. & Distl O. (2013) A genome-wide association study identifies risk loci to equine recurrent uveitis in German warmblood horses. *PLoS ONE* **8**, e71619.

Lampe V., Dierks C. & Distl O. (2009a) Refinement of a quantitative gene locus on equine chromosome 16 responsible for osteochondrosis in Hanoverian warmblood horses. *Animal* **3**, 1224–31.

Lampe V., Dierks C. & Distl O. (2009b) Refinement of a quantitative trait locus on equine chromosome 5 responsible for fetlock osteochondrosis in Hanoverian warmblood horses. *Animal Genetics* **40**, 553–5.

Lampe V., Dierks C., Komm K. & Distl O. (2009c) Identification of a new quantitative trait locus on equine chromosome 18 responsible for osteochondrosis in Hanoverian warmblood horses. *Journal of Animal Science* **87**, 3477–81.

Langlois B. & Blouin C. (2004) Practical efficiency of breeding value estimations based on annual earnings of horses for jumping, trotting, and galloping races in France. *Livestock Production Science* **87**, 99–107.

Langlois B. & Blouin C. (2007) Annual, career or single race records for breeding value estimation in race horses. *Livestock Science* **107**, 132–41.

Lanz S., Gerber V., Marti E., Rettmer H., Klukowska-Rotzler J., Gottstein B., Matthews J.B., Pirie S. & Hamza E. (2013) Effect of hay dust extract and cyathostomin antigen stimulation on cytokine expression by PBMC in horses with recurrent airway obstruction. *Veterinary Immunology and Immunopathology* **155**, 229–37.

Lassaline-Utter M., Gemensky-Metzler A.J., Scherrer N.M., Stoppini R., Latimer C.A., MacLaren N.E. & Myrna K.E. (2014) Corneal dystrophy in Friesian horses may represent a variant of pellucid marginal degeneration. *Veterinary Ophthalmology* **17** (Suppl 1), 186–94.

Lee J.R., Hong C.P., Moon J.W. *et al.* (2014) Genome-wide analysis of DNA methylation patterns in horse. *BMC Genomics* **15**, 598.

Leeb T., Vogl C., Zhu B. *et al.* (2006) A human-horse comparative map based on equine BAC end sequences. *Genomics* **87**, 772–6.

Leegwater P.A., Vos-Loohuis M., Ducro B.J. *et al.* (2016) Dwarfism with joint laxity in Friesian horses is associated with a splice site mutation in B4GALT7. *BMC Genomics* **17**, 839.

Leise B.S., Faleiros R.R., Watts M., Johnson P.J., Balck S.J. & Belknap J.K. (2010) Laminar inflammatory gene expression in the carbohydrate overload model of equine laminitis. *Equine Veterinary Journal* **43**, 54–61.

Lepeule J., Bareille N., Robert C., Ezanno P., Valette J.P., Jacquet S., Blanchard G., Denoix J.M. & Seegers H. (2009) Association of growth, feeding practices and exercise conditions with the prevalence of Developmental Orthopaedic Disease in limbs of French foals at weaning. *Preventive Veterinary Medicine* **89**, 167–77.

Lepeule J., Bareille N., Robert C., Valette J.P., Jacquet S., Blanchard G., Denoix J.M. & Seegers H. (2013) Association of growth, feeding practices and exercise conditions with the severity of the osteoarticular status of limbs in French foals. *The Veterinary Journal* **197**, 65–71.

Lewczuk D., Hecold M., Rusc A., Fraszczak M., Bereznowski A., Korwin-Kossakowska A., Kaminski S. & Szyda J. (2017) Single nucleotide polymorphisms associated with osteochondrosis dissecans in Warmblood horses at different stages of training. *Animal Production Science* **57**, 608–13.

Lewis S.L., Holl H.M., Streeter C., Posbergh C., Schanbacher B.J., Place N.J., Mallicote M.F., Long M.T. & Brooks S.A. (2017) Genomewide association study reveals a risk locus for equine metabolic syndrome in the Arabian horse. *Journal of Animal Science* **95**, 1071–9.

Librado P., Der Sarkissian C., Ermini L. *et al*. (2015) Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6889–97.

Librado P., Fages A., Gaunitz C. *et al*. (2016) The evolutionary origin and genetic makeup of domestic horses. *Genetics* **204**, 423–34.

Librado P., Gamba C., Gaunitz C. *et al*. (2017) Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442–5.

Lindgren G., Backstrom N., Swinburne J., Hellborg L., Einarsson A., Sandberg K., Cothran G., Vila C., Binns M. & Ellegren H. (2004) Limited number of patrilines in horse domestication. *Nature Genetics* **36**, 335–6.

Lindholm-Perry A.K., Kuehn L.A., Oliver W.T., Sexten A.K., Miles J.R., Rempel L.A., Cushman R.A. & Freetly H.C. (2013) Adipose and muscle tissue gene expression of two genes (NCAPG and LCORL) located in a chromosomal region associated with cattle feed intake and gain. *PLoS ONE* **8**, e80882.

Ling Y., Ma Y., Guan W., Cheng Y., Wang Y., Han J., Jin D., Mang L. & Mahmut H. (2010) Identification of Y chromosome genetic variations in Chinese indigenous horse breeds. *Journal of Heredity* **101**, 639–43.

Lippold S., Knapp M., Kuznetsova T. *et al*. (2011) Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nature Communications* **2**, 450.

Loftus J.P., Black S.J., Pettigrew A., Abrahamsen E.J. & Belknap J.K. (2007) Early laminar events involving endothelial activation in horses with black walnut- induced laminitis. *American Journal of Veterinary Research* **68**, 1205–11.

Lopes M.S., Diesterbeck U., da Camara Machado A. & Distl O. (2009) Fine mapping a quantitative trait locus on horse chromosome 2 associated with radiological signs of navicular disease in Hanoverian warmblood horses. *Animal Genetics* **40**, 955–7.

Lopes M.S., Diesterbeck U., Machado Ada C. & Distl O. (2010) Refinement of quantitative trait loci on equine chromosome 10 for radiological signs of navicular disease in Hanoverian warmblood horses. *Animal Genetics* **41**(Suppl 2), 36–40.

Lu J., Tang T., Tang H., Huang J., Shi S. & Wu C.I. (2006) The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends in Genetics* **22**, 126–31.

Ludwig E.K., Brandon Wiese R., Graham M.R., Tyler A.J., Settlage J.M., Werre S.R., Petersson-Wolfe C.S., Kanevsky-Mullarky I. & Dahlgren L.A. (2016) Serum and synovial fluid serum amyloid A response in equine models of synovitis and septic arthritis. *Veterinary Surgery* **45**, 859–67.

Lykkjen S., Dolvik N.I., McCue M.E., Rendahl A.K., Mickelson J.R. & Roed K.H. (2010) Genome-wide association analysis of osteochondrosis of the tibiotarsal joint in Norwegian Standardbred trotters. *Animal Genetics* **41**(Suppl 2), 111–20.

Lykkjen S., Roed K.H. & Dolvik N.I. (2012) Osteochondrosis and osteochondral fragments in Standardbred trotters: prevalence and relationships. *Equine Veterinary Journal* **44**, 332–8.

Mach N., Plancade S., Pacholewska A. *et al*. (2016) Integrated mRNA and miRNA expression profiling in blood reveals candidate biomarkers associated with endurance exercise in the horse. *Scientific Reports* **6**, 22932.

Mack M., Kowalski E., Grahn R., Bras D., Penedo M.C.T. & Bellone R. (2017) Two variants in SLC24A5 are associated with "tiger-eye" iris pigmentation in Puerto Rican Paso Fino horses. *G3 (Bethesda)*, **7**, 2799–806.

Makvandi-Nejad S., Hoffman G.E., Allen J.J. *et al*. (2012) Four loci explain 83% of size variation in the horse. *PLoS ONE* **7**, e39929.

Mansour T.A., Scott E.Y., Finno C.J., Bellone R.R., Mienaltowski M.J., Penedo M.C., Ross P.J., Valberg S.J., Murray J.D. & Brown C.T. (2017) Tissue resolved, gene structure refined equine transcriptome. *BMC Genomics* **18**, 103.

Marchiori C.M., Pereira G.L., Maiorano A.M., Rogatto G.M., Assoni A.D., Silva J.A.I.I.V., Chardulo L.A.L. & Curi R.A. (2019) Linkage disequilibrium and population structure characterization in the cutting and racing lines of Quarter Horses bred in Brazil. *Livestock Science* **219**, 45–51.

Marti E., Gerber H., Essich G., Oulehla J. & Lazary S. (1991) The genetic basis of equine allergic diseases. 1. Chronic hypersensitivity bronchitis. *Equine Veterinary Journal* **23**, 457–60.

Mason V.C., Schaefer R.J., McCue M.E., Leeb T. & Gerber V. (2018) eQTL discovery and their association with severe equine asthma in European Warmblood horses. *BMC Genomics* **19**, 581.

Maurano M.T., Humbert R., Rynes E. *et al*. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–5.

McCoy A.M. (2015) Animal models of osteoarthritis: comparisons and key considerations. *Veterinary Pathology* **52**, 803–18.

McCoy A.M., Beeson S.K., Splan R.K., Lykkjen S., Ralston S.L., Mickelson J.R. & McCue M.E. (2016) Identification and validation of risk loci for osteochondrosis in standardbreds. *BMC Genomics* **17**, 41.

McCoy A.M., Norton E.M., Kemper A.M., Beeson S.K., Mickelson J.R. & McCue M.E. (2019) SNP-based heritability and genetic architecture of tarsal osteochondrosis in North American Standardbred horses. *Animal Genetics* **50**, 78–81.

McCue M.E., Bannasch D.L., Petersen J.L. *et al*. (2012) A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genetics* **8**, e1002451.

McGivney C.L., Gough K.F., McGivney B.A., Farries G., Hill E.W. & Katz L.M. (2019) Exploratory factor analysis of signalment and conformational measurements in Thoroughbred horses with and without recurrent laryngeal neuropathy. *Equine Veterinary Journal* **51**, 179–84.

Meira C.T., Curi R.A., Farah M.M., de Oliveira H.N., Beltran N.A.R., Silva J.A.I.V. & da Mota M.D.S. (2014a) Prospection of genomic regions divergently selected in racing line of Quarter Horses in relation to cutting line. *Animal* **8**, 1754–64.

Meira C.T., Farah M.M., Fortes M.R.S., Moore S.S., Pereira G.L., Silva J.A.V., da Mota M.D.S. & Curi R.A. (2014b) A genome-wide association study for morphometric traits in quarter horse. *Journal of Equine Veterinary Science* **34**, 1028–31.

Meira C.T., Fortes M.R.S., Farah M.M., Porto-Neto L.R., Kelly M., Moore S.S., Pereira G.L., Chardulo L.A.L. & Curi R.A. (2014c) Speed index in the racing quarter horse: a genome-wide association study. *Journal of Equine Veterinary Science* **34**, 1263–8.

Metzger J., Ohnesorge B. & Distl O. (2012) Genome-wide linkage and association analysis identifies major gene loci for guttural pouch tympany in Arabian and German warmblood horses. *PLoS ONE* **7**, e41640.

Metzger J., Schrimpf R., Philipp U. & Distl O. (2013) Expression levels of LCORL are associated with body size in horses. *PLoS ONE* **8**, e56497.

Metzger J., Tonda R., Beltran S., Agueda L., Gut M. & Distl O. (2014) Next generation sequencing gives an insight into the characteristics of highly selected breeds versus non-breed horses in the course of domestication. *BMC Genomics* **15**, 562.

Metzger J., Karwath M., Tonda R., Beltran S., Agueda L., Gut M., Gut I.G. & Distl O. (2015) Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* **16**, 764.

Metzger J., Gast A.C., Schrimpf R., Rau J., Eikelberg D., Beineke A., Hellige M. & Distl O. (2017) Whole-genome sequencing reveals a potential causal mutation for dwarfism in the Miniature Shetland pony. *Mammalian Genome* **28**, 143–51.

Metzger J., Rau J., Naccache F., Bas Conn L., Lindgren G. & Distl O. (2018) Genome data uncover four synergistic key regulators for extremely small body size in horses. *BMC Genomics* **19**, 492.

Meyer M., Arsuaga J.L., de Filippo C. *et al.* (2016) Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–7.

Miga K.H., Newton Y., Jain M., Altemose N., Willard H.F. & Kent W.J. (2014) Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Research* **24**, 697–707.

Milinovich G.J., Trott D.J., Burrell P.C., van Eps A.W., Thoefner M.B., Blackall L.L., Al Jassim R.A., Morton J.M. & Pollitt C.C. (2006) Changes in equine hindgut bacterial populations during oligofructose-induced laminitis. *Environmental Microbiology* **8**, 885–98.

Mirams M., Tatarczuch L., Ahmed Y.A., Pagel C.N., Jeffcott L.B., Davies H.M. & Mackie E.J. (2009) Altered gene expression in early osteochondrosis lesions. *Journal of Orthopaedic Research* **27**, 452–7.

Mirams M., Ayodele B.A., Tatarczuch L., Henson F.M., Pagel C.N. & Mackie E.J. (2016) Identification of novel osteochondrosis—associated genes. *Journal of Orthopaedic Research* **34**, 404–11.

Molina A., Valera M., Dos Santos R. & Rodero A. (1999) Genetic parameters of morphofunctional traits in Andalusian horse. *Livestock Production Science* **60**, 295–303.

Moon S., Lee J.W., Shin D., Shin K.Y., Kim J., Choi I.Y., Kim J. & Kim H. (2015) A genome-wide scan for selective sweeps in racing horses. *Asian-Australasian Journal of Animal Sciences* **28**, 1525–31.

Moyers B.T., Morrell P.L. & McKay J.K. (2018) Genetic costs of domestication and improvement. *Journal of Heredity* **109**, 103–16.

Naccache F., Metzger J. & Distl O. (2018) Genetic risk factors for osteochondrosis in various horse breeds. *Equine Veterinary Journal* **50**, 556–63.

Nanaei H.A., Mehrgardi A.A. & Esmailizadeh A. (2019) Comparative population genomics unveils candidate genes for athletic performance in Hanoverians. *Genome* **62**, 279–85.

Nergadze S.G., Belloni E., Piras F.M., Khoriauli L., Mazzagatti A., Vella F., Bensi M., Vitelli V., Giulotto E. & Raimondi E. (2014) Discovery and comparative analysis of a novel satellite, EC137, in horses and other equids. *Cytogenetic and Genome Research* **144**, 114–23.

Nergadze S.G., Piras F.M., Gamba R. *et al.* (2018) Birth, evolution, and transmission of satellite-free mammalian centromeric domains. *Genome Research* **28**, 789–99.

Nolte W., Thaller G. & Kuehn C. (2019) Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. *PLoS ONE* **14**, e0215913.

Norton E.M., Avila F., Schultz N.E., Mickelson J.R., Geor R.J. & McCue M.E. (2019a) Evaluation of an HMGA2 variant for pleiotropic effects on height and metabolic traits in ponies. *Journal of Veterinary Internal Medicine* **33**, 942–52.

Norton E.M., Schultz N.E., Rendahl A.K., Geor R.J., Mickelson J.R. & McCue M.E. (2019b) Heritability of metabolic traits associated with equine metabolic syndrome in Welsh ponies and Morgan horses. *Equine Veterinary Journal* **51**, 475–80.

O'Ferrall G.J.M. & Cunningham E.P. (1974) Heritability of racing performance in thoroughbred horses. *Livestock Production Science* **1**, 87–97.

Ohno S. (1972) So much "junk" DNA in our genome. *Brookhaven Symposium in Biology* **23**, 366–70.

Ojala M., Vanvleck L.D. & Quaas R.L. (1987) Factors influencing best annual racing time in Finnish horses. *Journal of Animal Science* **64**, 109–16.

Orlando L., Ginolhac A., Raghavan M. *et al.* (2011) True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Research* **21**, 1705–19.

Orlando L., Ginolhac A., Zhang G. *et al.* (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–8.

Orr N., Hill E.W., Gu J. *et al.* (2013) Genome-wide association study of osteochondrosis in the tarsocrural joint of Dutch Warmblood horses identifies susceptibility loci on chromosomes 3 and 10. *Animal Genetics* **44**, 408–12.

Outram A.K., Stear N.A., Bendrey R., Olsen S., Kasparov A., Zaibert V., Thorpe N. & Evershed R.P. (2009) The earliest horse harnessing and milking. *Science* **323**, 1332–5.

Pääbo S. (1985) Molecular cloning of ancient Egyptian mummy DNA. *Nature* **314**, 644–5.

Pacholewska A., Jagannathan V., Drogemuller M., Klukowska-Rotzler J., Lanz S., Hamza E., Dermitzakis E.T., Marti E., Leeb T. & Gerber V. (2015) Impaired cell cycle regulation in a natural equine model of asthma. *PLoS ONE* **10**, e0136103.

Park W., Kim J., Kim H.J. *et al.* (2014) Investigation of de novo unique differentially expressed genes related to evolution in exercise response during domestication in Thoroughbred race horses. *PLoS ONE* **9**, e91418.

Patterson L., Staiger E.A. & Brooks S.A. (2015) DMRT3 is associated with gait type in Mangalarga Marchador horses, but does not control gait ability. *Animal Genetics* **46**, 213–5.

Penedo M.C., Millon L.V., Bernoco D. *et al.* (2005) International Equine Gene Mapping Workshop Report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources. *Cytogenetic and Genome Research* **111**, 5–15.

Pereira G.L., de Matteis R., Meira C.T., Regitano L.C.A., Silva J.A.V., Chardulo L.A.L. & Curi R.A. (2016) Comparison of sequence variants in the PDK4 and COX412 genes between racing and cutting lines of quarter horses and associations with the speed index. *Journal of Equine Veterinary Science* **39**, 1–6.

Pereira G.L., Malheiros J.M., Ospina A.M.T., Chardulo L.A.L. & Curi R.A. (2019) Exome sequencing in genomic regions related to racing performance of Quarter Horses. *Journal of Applied Genetics* **60**, 79–86.

Petersen J.L., Mickelson J.R., Cothran E.G. *et al*. (2013a) Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS ONE* **8**, e54997.

Petersen J.L., Mickelson J.R., Rendahl A.K. *et al*. (2013b) Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genetics* **9**, e1003211.

Petersen J.L., Mickelson J.R., Cleary K.D. & McCue M.E. (2014a) The American Quarter Horse: population structure and relationship to the Thoroughbred. *Journal of Heredity* **105**, 148–62.

Petersen J.L., Valberg S.J., Mickelson J.R. & McCue M.E. (2014b) Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fiber type proportions. *Animal Genetics* **45**, 827–35.

Physick-Sheard P., Kraus M., Basrur P., McGurrin K., Kenney D. & Schenkel F. (2014) Breed predisposition and heritability of atrial fibrillation in the Standardbred horse: a retrospective case-control study. *Journal of Veterinary Cardiology* **16**, 173–84.

Pielberg G., Mikko S., Sandberg K. & Andersson L. (2005) Comparative linkage mapping of the Grey coat colour gene in horses. *Animal Genetics* **36**, 390–5.

Piras F.M., Nergadze S.G., Magnani E., Bertoni L., Attolini C., Khoriauli L., Raimondi E. & Giulotto E. (2010) Uncoupling of satellite DNA and centromeric function in the genus Equus. *PLoS Genetics* **6**, e1000845.

Pirie R.S. (2014) Recurrent airway obstruction: a review. *Equine Veterinary Journal* **46**, 276–88.

Promerova M., Andersson L.S., Juras R. *et al*. (2014) Worldwide frequency distribution of the 'Gait keeper' mutation in the DMRT3 gene. *Animal Genetics* **45**, 274–82.

Purgato S., Belloni E., Piras F.M. *et al*. (2015) Centromere sliding on a mammalian chromosome. *Chromosoma* **124**, 277–87.

Rafati N., Andersson L.S., Mikko S. *et al*. (2016) Large deletions at the SHOX locus in the pseudoautosomal region are associated with skeletal atavism in Shetland Ponies. *G3 (Bethesda)*, **6**, 2213–23.

Ramseyer A., Gaillard C., Burger D., Straub R., Jost U., Boog C., Marti E. & Gerber V. (2007) Effects of genetic and environmental factors on chronic lower airway disease in horses. *Journal of Veterinary Internal Medicine* **21**, 149–56.

Raudsepp T., Gustafson-Seabury A., Durkin K. *et al*. (2008) A 4,103 marker integrated physical and comparative map of the horse genome. *Cytogenetic and Genome Research* **122**, 28–36.

Raudsepp T., McCue M.E., Das P.J. *et al*. (2012) Genome-wide association study implicates testis-sperm specific FKBP6 as a susceptibility locus for impaired acrosome reaction in stallions. *PLoS Genetics* **8**, e1003139.

Raudsepp T., Das P.J. & Chowdhary B.P. (2013) Genomics of reproduction and fertility. In: *Equine Genomics* (Ed. by B.P. Chowdhary), pp. 199–216. John Wiley & Sons, Ames, IA.

Rebolledo-Mendez J., Hestand M.S., Coleman S.J., Zeng Z., Orlando L., MacLeod J.N. & Kalbfleisch T. (2015) Comparison of the equine reference sequence with its sanger source data and new illumina reads. *PLoS ONE* **10**, e0126852.

Ricard A. & Chanu I. (2001) Genetic parameters of eventing horse competition in France. *Genetics Selection Evolution* **33**, 175–90.

Ricard A. & FournetHanocq F. (1997) Analysis of factors affecting length of competitive life of jumping horses. *Genetics Selection Evolution* **29**, 251–67.

Ricard A., Robert C., Blouin C. *et al*. (2017) Endurance exercise ability in the horse: a trait with complex polygenic determinism. *Frontiers in Genetics* **8**, 89.

Riddick T.L., Duesterdieck-Zellmer K. & Semevolos S.A. (2012) Gene and protein expression of cartilage canal and osteochondral junction chondrocytes and full-thickness cartilage in early equine osteochondrosis. *The Veterinary Journal* **194**, 319–25.

Robinson N.E., Derksen F.J., Olszewski M.A. & Buechner-Maxwell V.A. (1996) The pathogenesis of chronic obstructive pulmonary disease of horses. *British Veterinary Journal* **152**, 283–306.

Rooney M.F., Hill E.W., Kelly V.P. & Porter R.K. (2018) The "speed gene" effect of myostatin arises in Thoroughbred horses due to a promoter proximal SINE insertion. *PLoS ONE* **13**, e0205664.

Rosengren Pielberg G., Golovko A., Sundstrom E. *et al*. (2008) A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nature Genetics* **40**, 1004–9.

Saatchi M., Schnabel R.D., Taylor J.F. & Garrick D.J. (2014) Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* **15**, 442.

Sadeghi R., Moradi-Shahrbabak M., Miraei Ashtiani S.R., Schlamp F., Cosgrove E.J. & Antczak D.F. (2019) Genetic diversity of Persian Arabian horses and their relationship to other native Iranian horse breeds. *Journal of Heredity* **110**, 173–82.

Sadek M.H., Al-Aboud A.Z. & Ashmawy A.A. (2006) Factor analysis of body measurements in Arabian horses. *Journal of Animal Breeding and Genetics* **123**, 369–77.

Santagostino M., Khoriauli L., Gamba R. *et al*. (2015) Genome-wide evolutionary and functional analysis of the Equine Repetitive Element 1: an insertion in the myostatin promoter affects gene expression. *BMC Genetics* **16**, 126.

Schaefer R.J., Schubert M., Bailey E. *et al*. (2017) Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics* **18**, 565.

Schnider D., Rieder S., Leeb T., Gerber V. & Neuditschko M. (2017) A genome-wide association study for equine recurrent airway obstruction in European Warmblood horses reveals a suggestive new quantitative trait locus on chromosome 13. *Animal Genetics* **48**, 691–3.

Schrimpf R., Metzger J., Martinsson G., Sieme H. & Distl O. (2015) Implication of FKBP6 for male fertility in horses. *Reproduction in Domestic Animals* **50**, 195–9.

Schrimpf R., Gottschalk M., Metzger J., Martinsson G., Sieme H. & Distl O. (2016) Screening of whole genome sequences identified high-impact variants for stallion fertility. *BMC Genomics* **17**, 288.

Schroder W., Klostermann A., Stock K.F. & Distl O. (2012) A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian warmblood horses. *Animal Genetics* **43**, 392–400.

Schubert M., Jonsson H., Chang D. *et al*. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E5661–9.

Schurink A., Wolc A., Ducro B.J., Frankena K., Garrick D.J., Dekkers J.C. & van Arendonk J.A. (2012) Genome-wide

association study of insect bite hypersensitivity in two horse populations in the Netherlands. *Genetics Selection Evolution* **44**, 31.

Seiero T., Mark T. & Jonsson L. (2016) Genetic parameters for longevity and informative value of early indicator traits in Danish show jumping horses. *Livestock Science* **184**, 126–33.

Seong H.S., Kim N.Y., Kim D.C., Hwang N.H., Son D.H., Shin J.S., Lee J.H., Chung W.H. & Choi J.W. (2019) Whole genome sequencing analysis of horse populations inhabiting the Korean Peninsula and Przewalski's horse. *Genes Genomics* **41**, 621–8.

Sevane N., Dunner S., Boado A. & Canon J. (2016) Candidate gene analysis of osteochondrosis in Spanish Purebred horses. *Animal Genetics* **47**, 570–8.

Sevane N., Dunner S., Boado A. & Canon J. (2017) Polymorphisms in ten candidate genes are associated with conformational and locomotive traits in Spanish Purebred horses. *Journal of Applied Genetics* **58**, 355–61.

Shakhsi-Niaei M., Klukowska-Rotzler J., Drogemuller C., Swinburne J., Ehrmann C., Saftic D., Ramseyer A., Gerber V., Dolf G. & Leeb T. (2012) Replication and fine-mapping of a QTL for recurrent airway obstruction in European Warmblood horses. *Animal Genetics* **43**, 627–31.

Shin D.H., Lee J.W., Park J.E., Choi I.Y., Oh H.S., Kim H.J. & Kim H. (2015) Multiple genes related to muscle identified through a joint analysis of a two-stage genome-wide association study for racing performance of 1,156 thoroughbreds. *Asian-Australasian Journal of Animal Sciences* **28**, 771–81.

Signer-Hasler H., Flury C., Haase B., Burger D., Simianer H., Leeb T. & Rieder S. (2012) A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS ONE* **7**, e37282.

Signer-Hasler H., Neuditschko M., Koch C., Froidevaux S., Flury C., Burger D., Leeb T. & Rieder S. (2014) A chromosomal region on ECA13 is associated with maxillary prognathism in horses. *PLoS ONE* **9**, e86607.

Sild E., Rooni K., Värv S., Røed K., Popov R., Kantanen J. & Viinlass H. (2019) Genetic diversity of Estonian horse breeds and their genetic affinity to northern European and some Asian breeds. *Livestock Science* **220**, 57–66.

Snelling W.M., Allan M.F., Keele J.W., Kuehn L.A., McDaneld T., Smith T.P., Sonstegard T.S., Thallman R.M. & Bennett G.L. (2010) Genome-wide association study of growth in crossbred beef cattle. *Journal of Animal Science* **88**, 837–48.

Sole M., Bartolome E., Sanchez M.J., Molina A. & Valera M. (2017) Predictability of adult Show Jumping ability from early information: alternative selection strategies in the Spanish Sport Horse population. *Livestock Science* **200**, 23–8.

Staiger E.A., Al Abri M.A., Pflug K.M., Kalla S.E., Ainsworth D.M., Miller D., Raudsepp T., Sutter N.B. & Brooks S.A. (2016a) Skeletal variation in Tennessee Walking Horses maps to the LCORL/NCAPG gene region. *Physiological Genomics* **48**, 325–35.

Staiger E.A., Tseng C.T., Miller D., Cassano J.M., Nasir L., Garrick D., Brooks S.A. & Antczak D.F. (2016b) Host genetic influence on papillomavirus-induced tumors in the horse. *International Journal of Cancer* **139**, 784–92.

Steelman S.M., Chowdhary B.P., Dowd S., Suchodolski J. & Janecka J.E. (2012) Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Veterinary Research* **8**, 231.

Stock K.F. & Distl O. (2007) Genetic correlations between performance traits and radiographic findings in the limbs of German Warmblood riding horses. *Journal of Animal Science* **85**, 31–41.

Stock K.F., Hamann H. & Distl O. (2005) Estimation of genetic parameters for the prevalence of osseous fragments in limb joints of Hanoverian Warmblood horses. *Journal of Animal Breeding and Genetics* **122**, 271–80.

Sturtevant A.H. (1920) On the inheritance of coat color in harness horses. *Biological Bulletin* **19**, 204–16.

Suontama M., Saastamoinen M.T. & Ojala M. (2009) Estimates of non-genetic effects and genetic parameters for body measures and subjectively scored traits in Finnhorse trotters. *Livestock Science* **124**, 205–9.

Swinburne J.E., Bogle H., Klukowska-Rotzler J., Drogemuller M., Leeb T., Temperton E., Dolf G. & Gerber V. (2009) A whole-genome scan for recurrent airway obstruction in Warmblood sport horses indicates two positional candidate regions. *Mammalian Genome* **20**, 504–15.

Szwaczkowski T., Greguła-Kania M., Stachurska A., Borowska A., Jaworski Z. & Gruszecki T.M. (2016) Inter- and intra-genetic diversity in the Polish Konik horse: implications for the conservation program. *Canadian Journal of Animal Science* **96**, 570–80.

Tadros E.M., Frank N. & Horohov D.W. (2013) Inflammatory cytokine gene expression in blood during the development of oligofructose-induced laminitis in horses. *Journal of Equine Veterinary Science* **33**, 802–8.

Tavernier A. (1990) Estimation of breeding value of jumping horses from their ranks. *Livestock Production Science* **26**, 277–90.

Terry R.B., Archer S., Brooks S., Bernoco D. & Bailey E. (2004) Assignment of the appaloosa coat colour gene (LP) to equine chromosome 1. *Animal Genetics* **35**, 134–7.

Tetens J., Widmann P., Kuhn C. & Thaller G. (2013) A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Animal Genetics* **44**, 467–71.

Teyssedre S., Dupuis M.C., Guerin G., Schibler L., Denoix J.M., Elsen J.M. & Ricard A. (2012) Genome-wide association studies for osteochondrosis in French Trotter horses. *Journal of Animal Science* **90**, 45–53.

Thiruvenkadan A.K., Kandasamy N. & Panneerselvam S. (2009a) Inheritance of racing performance of Thoroughbred horses. *Livestock Science* **121**, 308–26.

Thiruvenkadan A.K., Kandasamy N. & Panneerselvam S. (2009b) Inheritance of racing performance of trotter horses: an overview. *Livestock Science* **124**, 163–81.

Thomer A., Gottschalk M., Christmann A., Naccache F., Jung K., Hewicker-Trautwein M., Distl O. & Metzger J. (2018) An epistatic effect of KRT25 on SP6 is involved in curly coat in horses. *Scientific Reports* **8**, 6374.

Tozaki T., Miyake T., Kakoi H., Gawahara H., Sugita S., Hasegawa T., Ishida N., Hirota K. & Nakano Y. (2010) A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. *Animal Genetics* **41**(Suppl 2), 28–35.

Tozaki T., Kikuchi M., Kakoi H., Hirota K.I. & Nagata S.I. (2017) A genome-wide association study for body weight in Japanese Thoroughbred racehorses clarifies candidate regions on chromosomes 3, 9, 15, and 18. *Journal of Equine Science* **28**, 127–34.

Tozaki T., Kikuchi M., Kakoi H. *et al*. (2019) Genetic diversity and relationships among native Japanese horse breeds, the Japanese Thoroughbred and horses outside of Japan using genome-wide SNP data. *Animal Genetics* **50**, 449–59.

Treiber K.H., Kronfeld D.S., Hess T.M., Byrd B.M., Splan R.K. & Staniar W.B. (2006) Evaluation of genetic and metabolic predispositions and nutritional risk factors for pasture-associated laminitis in ponies. *Journal of the American Veterinary Medical Association* **228**, 1538–45.

Tsang H.G., Rashdan N.A., Whitelaw C.B., Corcoran B.M., Summers K.M. & MacRae V.E. (2016) Large animal models of cardiovascular disease. *Cell Biochemistry and Function* **34**, 113–32.

Tuggle C.K., Giuffra E., White S.N. *et al*. (2016) GO-FAANG meeting: a gathering on functional annotation of animal genomes. *Animal Genetics* **47**, 528–33.

Valberg S.J., Henry M.L., Perumbakkam S., Gardner K.L. & Finno C.J. (2018) An E321G MYH1 mutation is strongly associated with nonexertional rhabdomyolysis in Quarter Horses. *Journal of Veterinary Internal Medicine* **32**, 1718–25.

Vandenput S., Duvivier D.H., Votion D., Art T. & Lekeux P. (1998a) Environmental control to maintain stabled COPD horses in clinical remission: effects on pulmonary function. *Equine Veterinary Journal* **30**, 93–6.

Vandenput S., Votion D., Duvivier D.H., Van Erck E., Anciaux N., Art T. & Lekeux P. (1998b) Effect of a set stabled environmental control on pulmonary function and airway reactivity of COPD affected horses. *Veterinary Journal* **155**, 189–95.

Vander Heyden L., Lejeune J.P., Caudron I., Detilleux J., Sandersen C., Chavatte P., Paris J., Deliege B. & Serteyn D. (2013) Association of breeding conditions with prevalence of osteochondrosis in foals. *Veterinary Record* **172**, 68.

Velie B.D., Shrestha M., Franois L. *et al*. (2016) Using an inbred horse breed in a high density genome-wide scan for genetic risk factors of insect bite hypersensitivity (IBH). *PLoS ONE* **11**, e0152966.

Viklund A., Braam A., Nasholm A., Strandberg E. & Philipsson J. (2010) Genetic variation in competition traits at different ages and time periods and correlations with traits at field tests of 4-year-old Swedish Warmblood horses. *Animal* **4**, 682–91.

Vila C.L.J., Gotherstrom A., Marklund S., Sandberg K., Liden K., Wayne R.K. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291**, 474–7.

Viluma A., Mikko S., Hahn D., Skow L., Andersson G. & Bergstrom T.F. (2017) Genomic structure of the horse major histocompatibility complex class II region resolved using PacBio long-read sequencing technology. *Scientific Reports* **7**, 45518.

Wade C.M. (2013) Assembly and analysis of the equine genome sequence. In: *Equine Genomics* (Ed. by B.P. Chowdhary), pp. 103–12. Wiley-Blackwell, Ames, IA.

Wade C.M., Giulotto E., Sigurdsson S. *et al*. (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865–7.

Wallin L., Strandberg E. & Philipsson J. (2003) Genetic correlations between field test results of Swedish Warmblood Riding Horses as 4-year-olds and lifetime performance results in dressage and show jumping. *Livestock Production Science* **82**, 61–71.

Wallner B., Piumi F., Brem G., Muller M. & Achmann R. (2004) Isolation of Y chromosome-specific microsatellites in the horse and cross-species amplification in the genus Equus. *Journal of Heredity* **95**, 158–64.

Wallner B., Vogl C., Shukla P., Burgstaller J.P., Druml T. & Brem G. (2013) Identification of genetic variation on the horse y chromosome and the tracing of male founder lineages in modern breeds. *PLoS ONE* **8**, e60015.

Wallner B., Palmieri N., Vogl C. *et al*. (2017) Y chromosome uncovers the recent oriental origin of modern stallions. *Current Biology* **27**, 2029–35 e5.

Wang E.T., Sandberg R., Luo S., Khrebtukova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P. & Burge C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–6.

Wang W., Wang S., Hou C. *et al*. (2014) Genome-wide detection of copy number variations among diverse horse breeds by array CGH. *PLoS ONE* **9**, e86860.

Warmuth V., Eriksson A., Bower M.A. *et al*. (2011) European domestic horses originated in two holocene refugia. *PLoS ONE* **6**, e18194.

Warmuth V., Eriksson A., Bower M.A. *et al*. (2012) Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8202–6.

Weikard R., Altmaier E., Suhre K., Weinberger K.M., Hammon H.M., Albrecht E., Setoguchi K., Takasuga A. & Kuhn C. (2010) Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on *Bos taurus* growth and lipid deposition. *Physiological Genomics* **42A**, 79–88.

Welsh C.E., Lewis T.W., Blott S.C., Mellor D.J., Lam K.H., Stewart B.D. & Parkin T.D. (2013) Preliminary genetic analyses of important musculoskeletal conditions of Thoroughbred racehorses in Hong Kong. *Veterinary Journal* **198**, 611–5.

van der Weyden L., Patton E.E., Wood G.A., Foote A.K., Brenn T., Arends M.J. & Adams D.J. (2016) Cross-species models of human melanoma. *Journal of Pathology* **238**, 152–65.

Williamson S.A. & Beilharz R.G. (1996) Heritabilities of racing performance in Thoroughbreds: a study of Australian data. *Journal of Animal Breeding and Genetics-Zeitschrift Fur Tierzuchtung Und Zuchtungsbiologie* **113**, 505–24.

Witkowski L., Cywinska A., Paschalis-Trela K., Crisman M. & Kita J. (2016) Multiple etiologies of equine recurrent uveitis—a natural model for human autoimmune uveitis: a brief review. *Comparative Immunology, Microbiology and Infectious Diseases* **44**, 14–20.

Wittwer C., Hamann H., Rosenberger E. & Distl O. (2006) Prevalence of osteochondrosis in the limb joints of South German Coldblood horses. *Journal of Veterinary Medicine. A, Physiology, Pathology, Clinical Medicine* **53**, 531–9.

Wittwer C., Lohring K., Drogemuller C., Hamann H., Rosenberger E. & Distl O. (2007) Mapping quantitative trait loci for osteochondrosis in fetlock and hock joints and palmar/plantar osseus fragments in fetlock joints of South German Coldblood horses. *Animal Genetics* **38**, 350–7.

Woods P.S.A., Robinson N.E., Swanson M.C., Reed C.E., Broadstone R.V. & Derksen F.J. (1993) Airborne dust and aeroallergen concentration in a horse stable under 2 different management-systems. *Equine Veterinary Journal* **25**, 208–13.

Wutke S., Sandoval-Castellanos E., Benecke N. *et al*. (2018) Decline of genetic diversity in ancient domestic stallions in Europe. *Science Advances* **4**, eaap9691.

Yue F., Cheng Y., Breschi A. *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–64.

Zechner P., Zohman F., Solkner J., Bodo I., Habe F., Marti E. & Brem G. (2001) Morphological description of the Lipizzan horse population. *Livestock Production Science* **69**, 163–77.

Zhang C., Ni P., Ahmad H.I. *et al.* (2018) Detecting the population structure and scanning for signatures of selection in horses (*Equus caballus*) from whole-genome sequencing data. *Evolutionary Bioinformatics Online* **14**, 1176934318775106.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Genetic variants identified for traits influencing pigmentation.

Table S2 Genetics variants underlying disease and performance traits in the horse.